DOCUMENT RESUME

ED 097 640 CS 001 381

AUTHOR Siegel, Arthur I.; And Others
TITLE Increasing and Evaluating the Readability of Air Force Written Materials.
INSTITUTION Air Force Human Resources Lab., Lowry AFB, Colo. Technical Training Div.
REPORT NO AFHRL-TR-74-28
PUB DATE Aug 74
NOTE 87p.; See related document CS 001 383

EDRS PRICE MF-$0.75 HC-$4.20 PLUS POSTAGE
DESCRIPTORS Armed Forces; *Readability; Reading; *Reading Comprehension; Reading Instruction; Reading Level; *Reading Materials; Reading Research
IDENTIFIERS *Air Force

ABSTRACT

This report describes how to apply techniques that have been used in measuring the readability/comprehensibility and reading level of textural materials. A review of available readability/comprehensibility measurement techniques and of previous experimental studies into methods for developing such techniques is presented. As a result of this review, techniques for reading level determination and methods for increasing the readability/comprehensibility of textual materials are discussed. Instructions are provided in a step-by-step fashion for determining the reading level of written material and for presenting subject matter material through methods other than prose. In addition, procedures for simplifying written material and for determining the effectiveness of written material are described. (WR)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

AFHRL-TR-74-28

BEST COPY AVAILABLE

INCREASING AND EVALUATING THE READABILITY OF AIR FORCE WRITTEN MATERIALS

By

Arthur I. Siegel
Philip J. Federman
Applied Psychological Services, Inc.
Wayne, Pennsylvania 19087

James R. Burkett

TECHNICAL TRAINING DIVISION
Lowry Air Force Base, Colorado 80230

August 1974

Approved for public release; distribution unlimited.

AIR FORCE HUMAN RESOURCES LABORATORY

BEST COPY AVAILABLE
ED 097640

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235

NOTICE

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This final report was submitted by Applied Psychological Services, Inc., Wayne, Pennsylvania 19087, under contract F41609-72-C-0011, project 1121, with the Technical Training Division, Air Force Human Resources Laboratory (AFSC), Lowry Air Force Base, Colorado 80230. Dr. James R. Burkett, Instructional Technology Branch, was the contract monitor.

This report has been reviewed and cleared for open publication and/or public release by the appropriate Office of Information (OI) in accordance with AFR 190-17 and DoDD 5230.9. There is no objection to unlimited distribution of this report to the public at large, or by DDC to the National Technical Information Service (NTIS).

This report has been reviewed and is approved.

MARTY R. ROCKWAY, Technical Director
Technical Training Division

Approved for publication.

HAROLD E. FISCHER, Colonel, USAF
Commander

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFHRL-TR-74-28	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) INCREASING AND EVALUATING THE READABILITY OF AIR FORCE WRITTEN MATERIALS		5. TYPE OF REPORT & PERIOD COVERED Final 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Arthur I. Siegel Philip J. Federman James R. Burkett		8. CONTRACT OR GRANT NUMBER(s) F41609-72-C-0011
9. PERFORMING ORGANIZATION NAME AND ADDRESS Applied Psychological Services, Inc. Wayne, Pennsylvania 19087		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 62703F 11210403
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Human Resources Laboratory (AFSC) Brooks Air Force Base, Texas 78235		12. REPORT DATE August 1974 13. NUMBER OF PAGES 90
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Technical Training Division Air Force Human Resources Laboratory Lowry Air Force Base, Colorado 80230		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

readability
technical writing
reading levels
reading comprehension
textual materials

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes how to apply techniques that have been used in measuring the readability/comprehensibility and reading level of textual materials. Instructions are provided, in a step-by-step fashion, for determining the reading level of written material and for presenting subject matter material through methods other than prose. In addition, procedures for simplifying written material are presented. Experimental procedures, to be used in determining the effectiveness of written material, are described. Finally, multisensory presentation of Air Force training material is examined.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SUMMARY

Problem

This manual presents techniques and considerations for reducing the reading level of Air Force training materials so as to improve the capability of these materials to transfer information to the student.

Approach

A review of available readability/comprehensibility measurement techniques and previous experimental studies into methods for developing such techniques is presented. As the result of this review, techniques for reading level determination and methods for increasing the readability/comprehensibility of textual materials are discussed. The application of the techniques is demonstrated, and practical considerations, including cost/effectiveness determination, for the Air Force are considered. Methods for presenting subject matter material, other than through written prose, are described. In order to determine the effectiveness of a modified technique, performance followups are suggested. Various experimental techniques and testing procedures for this purpose are discussed.

Results

Several techniques for measuring readability are described in detail, and methods for their use and application are fully delineated. The readability techniques included are: cloze, Flesch, FORCAST, SMOG, and the Automated Readability Index (ARI). A quantitative approach for determining the cost/effectiveness of a subject matter presentation technique is developed and described. Instructions are provided for presenting material through auditory cassettes and for preparing illustrations of various types. Techniques for simplifying material are explained.

Conclusions

Application of the various methods and techniques described will reduce training time and costs and will serve to increase the level of comprehension by students who are exposed to various training materials.

TABLE OF CONTENTS

Page

I. INTRODUCTION 7

Techniques for Measuring Readability 7
Early (1920s) Studies into Analytic Methods 8
Studies During the 1930s 10
Studies of the 1940s and Early 1950s 12
Recent Studies 14
The Flesch Method 17
The Cloze Procedure 18
The Dale and Chall Method 20
Reading Level Determination 20

II. APPLICATIONS FOR INCREASING RELIABILITY 23

Gains to be Anticipated 23
Time for Comprehension 24
Logical and Precise Writing 24
Word and Sentence Simplification 27
Repetition-Redundancy 31

Textual Supplementation 31

Auditory Supplementation 31
Illustration Supplementation 33
Flowcharts and Tables 35
Graphs 39
Schematic Diagrams 43

III. "NONANALYTIC" METHODS FOR EVALUATING TEXTUAL METHODS 49

Design of Materials Evaluations 49
Single Group Methods 50
Control Group Studies 50
Factorial Design 51
Conduct of Nonanalytic Studies 53
Characteristics of Useful Measure of Performance 54

Page

Examples of Dependent Variables 55

Written Tests 56
Criterion and Norm Referenced Testing 56
Performance Tests 57
The Evaluative Milieu 58
Additional Reading List 59

REFERENCES 60

APPENDIX A How to Perform a Cloze Analysis 63

APPENDIX B Examples of Before and After Material and Automated Reading Index (ARI) and Grade Level (GL) Analyses 68

APPENDIX C How to Perform a Flesch Analysis 71

APPENDIX D How to Perform the FORCAST Analysis 77

APPENDIX E How to Perform the SMOG Analysis 78

APPENDIX F Glossary of Readability Measures 81

LIST OF ILLUSTRATIONS

Figure		Page
1	Sample of instructions in prose	35
2	Example of flowchart to represent instructions of Exhibit I	36
3	Example of tabular presentation of instructions	38
4	Weight as a function of height for men and women between the ages of 18 and 64	41
5	Annual mean income of males (25 years and older) in 1956 and 1965 as a function of years of education	42
6	Percentage of United States' population completing various years of schooling, 1966	44
7	Example of schematic incorporating comprehensibility principles	46

I. INTRODUCTION

The mismatch between the reading grade level of military personnel and the readability level of the material they read has been found to be quite vast in some instances (i. e., four to eight grade levels). This manual presents techniques which can be used to improve and measure the readability/intelligibility of Air Force reading materials and, accordingly, reduce the gap between the reading grade levels of the readers and the material they read.

This manual is intended to serve as a source document for use by Air Force training and development personnel, as well as others who prepare reading material for the Air Force. The techniques presented are described historically, evaluated, and demonstrated. Methods for evaluating the comprehensibility of reading material are discussed. Finally, several practical supplements are presented. These contain information regarding auditory, pictorial, and multisensory presentation of information. Such considerations are important to the preparation of materials used in training some types of personnel having limited reading skills.

Techniques for Measuring Readability

The readability of written material is linked to the ease of understanding the material and to the speed at which the printed matter can be read. A large number of methods are available for assessing readability. In this manual, we group them under: (1) analytic, and (2) nonanalytic methods. Analytic methods include any method which involves isolation of the various characteristics of the textual material itself. Nonanalytic methods are based on some form of measurement of what has been learned by the reader from the text.

Most of the analytic methods are based on some formula which considers the structural attributes (e. g., sentence length, word length, sentence complexity, use of familiar words) of the written material. Such formulas provide a quantitative index of the readability of the material. Under analytic methods, we also include the "cloze" technique (Appendix A).

Although the readability formulas are practical in many situations, they also suffer from certain drawbacks that can be avoided through the use of the nonanalytic approaches. Specifically, while the analytic methods provide an index of various attributes of written material (such as difficulty, reading grade level, interest), they tell us little about how much is learned from reading the text. The nonanalytic methods provide such information.

Early (1920s) Studies into Analytic Methods

Initial analytic efforts to measure readability objectively were sparked by the difficulty experienced by teachers when they attempted to select the most appropriate texts for their classes. Science teachers, in particular, found that the textbooks intended for classroom use contained a vocabulary level which was beyond that of the students. Consequently, objective methods were developed for determining the vocabulary difficulty of textbooks, so that those books with exceptionally difficult vocabulary could be avoided.

One early technique was developed in 1923 by Lively and Pressey. They developed an index number to reflect vocabulary difficulty. In order to obtain the index, a sample of 1,000 words, evenly distributed throughout a text, was taken. Then reference was made to a list of the 10,000 most commonly occurring words in the English language. This list was compiled by Thorndike in 1921. Values were assigned to each word in the sample, such that if the word appeared in the most common thousand it received an index value of 10; those words in the second thousand received a value of 9, and so on down through the last thousand-word block in the list. Words not appearing on the list were assigned an index value of zero. The difficulty of the vocabulary of the text under consideration was computed by obtaining the median of the index values assigned to each word. The resulting vocabulary difficulty indices agreed substantially with the judged difficulty of a wide variety of reading selections (from second grade reading to college level). However, such a readability determination technique possesses a number of drawbacks. These include considerations of: (1) the appropriateness of the 10,000 word reference list, (2) the small size of the sample (samples of more than 1,000 words from an entire book may be more appropriate), (3) prior experience with the subject matter area (readability may be easy, despite difficult vocabulary, because of familiarity with the subject matter, or readability may be difficult, despite easy vocabulary, because of a lack of familiarity with the

subject matter), (4) difficulty of the concepts, (5) the reader's general intelligence, and (6) his interest in the subject matter.

Washburne and Vogel, working in the Illinois school system, generated norms for use with the Lively and Pressey method. These norms were based on calculated Lively and Pressey index values for 700 books that were reported as having been read and liked during the preceding year by at least 25 students. They also determined the difficulty level for 150 of the 700 books. To do this, books read by lower grade level students and by higher grade level students were separated. Several factors were found that related to the criterion of grade level (these factors could distinguish those books read by lower grade pupils from those read by higher grade pupils). These factors correlated significantly with the difficulty of the textual material. Four of these factors were set into a formula for predicting readability (Y). The four factors were: number of different words appearing (x_1), number of occurrences of prepositions (x_2), number of words not on Thorndike's list of 10,000 (x_3), and a number of simple sentences in a 75 sentence sample (x_4). A regression equation was derived for determining the difficulty of those books reported as being read and liked. The equation for predicting grade levels as measured by the Standard Paragraph Meaning Test (Y) is:

$$Y = .085x_1 + .101x_2 + .604x_3 - .411x_4 + 17.43$$

A drawback inherent in this study is that the reading material sample included only those books which were read and liked. Accordingly, the sample is confounded by the evaluation of the books by the students. However, the Washburne and Vogel technique is significant historically because it: (1) analyzed the effects of structural factors in the text, (2) employed an objective criterion (textual difficulty) instead of qualitative judgment, and (3) described readability in terms of school grade levels rather than in terms of a relative measure.

The two readability techniques, discussed previously, that of Lively and Pressey and that of Washburne and Vogel, are typical of the early measures which emphasized readability measurement as a function of vocabulary uniqueness with reference to Thorndike's word list, and which used relatively crude criteria of the difficulty of the text.

Studies During the 1930s

The readability studies undertaken during the 1930s were largely characterized by: (1) emphasis on evaluation of a greater number of readability factors, (2) reduced emphasis on word frequency lists (e.g., Thorndike's 10,000 most commonly occurring words in the English language), and (3) a trend to nonanalytic methods, i.e., concern over more objective readability criteria such as comprehension test scores.

For example, Ojemann studied the use of comprehension test scores as the criterion of readability. He administered 16 passages, each of approximately 500 words, to adults and then tested their comprehension of each passage. The passages were arranged in order of difficulty and their contents were analyzed for eight sentence factors, six vocabulary factors, and three qualitative factors. Instead of a formula for determining readability, Ojemann presented his 16 listed selections, along with the values he determined for the quantitative aspects (i.e., the vocabulary and sentence factors), and their respective difficulty levels. This type of presentation was intended to allow other analysts to evaluate the qualitative factors in new passages, in addition to comparing the new passages with his tested passages on the quantitative factors. While such a technique is cumbersome and time consuming, Ojemann's study is noteworthy because it was the first to: employ adult subjects, use comprehension test scores as a criterion of readability, and demonstrate the value of qualitative, nonstatistical factors in determining readability.

However, this period was not devoid of some continued emphasis on the development of formulas. For example, Dale and Tyler asked adults of low reading ability (third to fifth grade level) to take a comprehension test after reading health education articles that appeared in newspapers, magazines, and books. Adequate test scores were not obtained. Dale and Tyler rewrote the passages, making them much simpler to read. They found that using very basic vocabulary, an informal style characterized by a conversational manner and anecdotes, and never digressing from the main topic increased the readability of the articles. Their analysis, based on the readministration of the rewritten passages, resulted in several factors correlating significantly with comprehension. Three of the factors were included in a regression equation for predicting comprehension: the number of technical words (x_1), the number of nontechnical words not known to 90 per cent of sixth grade pupils, as determined in a study performed by Dale (x_2), and the number of indeterminate clauses (x_3). The equation used to predict comprehension (Y) is:

$$Y = -9.4x_1 - 0.4x_2 + 2.2x_3 + 114.4$$

One study which fully exemplifies the trend in formula extension during this period is that of Gray and Leary. They listed over 80 factors related to readability. Through various processes of elimination, Gray and Leary were able to develop regression equations based on only a few factors. The most frequently used of their formulas was one consisting of five factors. This formula is

$$Y = .01029x_2 + .C09012x_5 - .02094x_6 - .03313x_7 - .01485x_8 + 3.774$$

where:

Y = predicted average comprehension score, along a scale of + 4 to -4

x_2 = average number of difficult words (words not appearing on Dale's list of 769 easy words)

x_5 = average number first, second, and third person pronouns

x_6 = average sentence length in words

x_7 = average percentage of different words

x_8 = average number of prepositional phrases

Since this formula is most applicable to low ability readers, it lacks broad utility.

Studies of the 1940s and Early 1950s

The trend in readability research during the 1940-1950 decade tended toward more efficient and more easily applied formulas. Lorge modified the formulas developed by Gray and Leary and obtained higher correlations with his criterion of reading passages than did Gray and Leary. Lorge attributed this increased correlation to his use of a greater and more adequately standardized set of criterion materials. Lorge's criterion materials were the Standard Test Lessons in Reading of McCall and Crabbs. The Standard Test Lessons in Reading consist of 376 reading passages containing norms related to grade levels. The formula Lorge used included two of Gray and Leary's factors, x_6 (average sentence length in words) and x_8 (number of prepositional phrases per hundred words), in addition to the factor x_9 (ratio of hard words to total number of words, hard words being those that do not appear on Dale's list of 769 easy words). The final readability formula was:

$$C_{50} = .10x_2 + .06x_6 + .10x_8 + 1.99$$

where:

C_{50} = reading grade level of individuals answering 50 per cent of the McCall and Crabbs questions correctly

Flesch was dissatisfied with the then current readability formulas because of the poor relationship found between the rankings of magazine articles (as determined by the formulas) and the rankings of the educational levels of the magazines' readers. He thought that the misranking was due to an emphasis on vocabulary and insufficient stress on other factors. He suggested three factors that should highly influence readability for adults: (1) sentence length, (2) abstracts, as determined through the number of affixes (i.e., prefixes and suffixes), and (3) number of personal reference words (i.e., pronouns, names, and words relating to people such as aunt, baby, etc.) Flesch's initial formula underwent a revision several years after it was first developed. This revised formulation which yields a reading ease and a human interest index has become widely used for readability analytic purposes. This revised formulation for measuring readability will be discussed in another section of this chapter.

Dale and Chall employed Flesch's original formula to evaluate the readability (for the average adult) of educational material published by the National Tuberculosis Association. They sought to develop their own measure of readability because they were dissatisfied with two aspects of the Flesch formula. To them, counting of affixes (prefixes and suffixes) was not completely manageable. Dale and Chall found that different raters, counting affixes in the same passage, came up with different numbers. However, they did agree with Flesch that a measure of abstractness was of value and had a place in a readability formula. Dale and Chall's second objection to the Flesch formula related to the use of personal words. They felt that Flesch's interest formula was overly simplified and that the definition of personal words (for purposes of readability) should be more restrictive. Dale and Chall's formula for measuring readability will also be discussed more fully in a subsequent section of this chapter, since it is also a popular formula for measuring readability today.

Another readability formula produced in this era was that of Powers, Sumner, and Kearl. Using the revised passages of McCall and Crabb's Standard Test Lessons in Reading (which was updated so that 60 passages dealt with then modern topics), Powers, Sumner, and Kearl recomputed Flesch's formulas, as well as others that had been developed earlier. They intended to produce formulas which would account for changes in reading abilities over the period between the two test editions of the Standard Test Lessons in Reading (some 24 years) and to facilitate comparisons of the formulas they recomputed. The researchers claimed that the revised formulas provided estimates more consistent with one another than did the original Flesch and Dale-Chall formulas.

McElroy's entry into the field of readability resulted in a measure referred to as the "fog count." The fog count is also related to the Flesch formulas in that it too counts syllables. Weights are assigned to words such that one and two syllable words receive a value of 1; a value of 3 is given to each remaining word. The sum of these weights equals the fog count. The reading grade level was determined from the fog count in the following manner: divide by 2 for sums of 20 and over; subtract 2 from sums under 20, and then divide by 2.

Recent Studies

The trends continued in the 1940-early 1950 period. However, there has recently been some emphasis on linguistic theoretic approaches to comprehensibility. Spache, in 1953, developed a formula to predict primary grade level (grades 1-3) of textual material on the basis of sentence length and percentage of words not appearing on the Dale list of 769 common words. The formula is:

$$\text{Grade level} = .141x_1 + .086x_2 + .839$$

where:

x_1 = average sentence length in words

x_2 = percentage of words not appearing on Dale's list of 769 common words

Wheeler and Smith, in 1954, developed a formula which equated grade level to 10 times the product of the mean length of units (sentences) in words and the percentage of multisyllable words. The value determined by the multiplication of these factors was located in a table and grade levels 1 through 4 could then be read from the table.

Tribe, in 1956, used the revised form of the McCall and Crabbs Standard Test Lessons in Reading as the criterion of readability. The grade level score of pupils, in grades 2 through 8, who could answer 50 per cent of the reading test questions correctly is:

$$.0719x_1 + .1043x_5 + 2.9347$$

where:

x_1 = average sentence length in words

x_5 = percentage (times 100) of words not appearing on a standard word list

In 1954 and 1958, Flesch introduced three alternate readability indices. These have not become highly popular, possibly because all of the indices are based on measures of word counts. The counted values are converted to arbitrary scales by referring to conversion tables. Flesch's "r" score is an index of realism, based on the number of references to specific human beings, their attributes or possessions, locations, objects numbered or named, dates, times, and colors. His "e" score is an index of energy, based on indications of voice communications, such as inflection. Finally, the "formality-popularity" scale is based on the total number of capitalized, underlined or italicized words, numbers (not spelled out), punctuation marks, symbols (#, $), beginnings of paragraphs, and endings of paragraphs.

Coleman, in 1965, developed four formulas using from one to four variables to predict readability. The readability criterion was the mean "cloze" score on the passage achieved by a sample of college students. The "cloze" score is the percentage of deleted words in a passage filled in correctly by a reader. The "cloze" technique will be discussed more fully below. The simplest of Coleman's formulas, which correlates very highly with the involved formulas, is:

Predicted cloze score = 1.29 (percentage of one syllable words) - 38.45

McLaughlin, In 1969, found that the number of syllables in 100 words is equal to three times the number of words of over two syllables, plus 112. A very close approximation to his formula, referred to as the SMOG grade, is:

SMOG Grade = 3 + square root of the polysyllable count in 30 sentences

where:

polysyllable = number of words with three or more syllables

A formula for measuring readability was developed at HUMRRO in 1972. The measure is referred to as FORCAST and was developed for measuring the readability of Army technical literature. The developers of this formula considered existing formulas inappropriate for Army technical material because extant formulas were most often based on school students and on general text material. Cloze scores were used as the criterion of readability. It was noted that cloze consistently yielded very high correlations with multiple choice tests and other more subjective measures of comprehension and difficulty. Prior research indicated that if a cloze score of 35 per cent was achieved by a reader for a given passage, then he might be expected to achieve a score of 70 per cent on a multiple-choice test of the same reading passage. Hence, cloze scores were demonstrated to be good indicators of comprehension and reading achievement level. The FORCAST formula for predicting reading grade level (Y) is:

$$Y = 20 - (\text{number of one syllable words occurring in a 150 word sample}/10)$$

Smith and Senter's readability analytic technique was unique because it was based on data which could be collected from mechanical counters that are installed on an IBM Selectric typewriter. The technique allows measurement of readability at rough draft typing speed. The mechanical counters record the number of key strokes, spaces, and sentences. The mean number of words per sentence (number of spaces divided by the number of sentences, w/s) and the mean length of words (number of strokes divided by the number of spaces, s/w) may be computed. The regression equation which predicts grade level (GL) from the above ratios and which was based on graded school texts is:

$$GL = 0.50\,(w/s) + 4.71\,(s/w) - 21.43$$

This formula may be simplified to:

$$GL = w/s + 9(s/w)$$

As such, it is referred to as the Automated Readability Index (ARI). In addition to the speed of data collection that it affords, the advantages of the ARI technique are: (1) its objectivity, (2) its reliability, and (3) the ease with which it may be incorporated into modern computerized typesetting machinery.

The Flesch Method

The most widely employed Flesch formulation rests on two indices--reading ease and human interest. The reading ease (RE) formula is:

$$RE = 206.835 - .846\,WL - 1.015\,SL$$

or:

$$RE = 206.835 - [1.015(SL) + .846(WL)]$$

where:

WL = word length measured as syllables per 100 words

SL = average sentence length in words for a given sample

The human interest (HI) index is determined from the formula:

$$HI = 3.635\,PW + .314\,PS$$

where:

PW = average percentage of personal words

PS = percentage of sentences addressed to the reader (including exclamations and grammatically incomplete sentences whose meaning is interpreted from the context of the passages)

This index tests the conversational style of the passage. The conjecture behind the human interest factor is that prose written in a conversational style will be more readable than that which is nonconversationally oriented.

These Flesch formulas provide index values which range from 0 to 100, and have become the most widely applied of any which have been developed. The wide use of the formulas is due, in part, to their ease of application.

A simplification of the Flesch reading ease formula was introduced by Farr, Jenkins, and Patterson in 1951. They suggested that Flesch's formula use the number of one syllable words, thereby allowing a more rapid test, instead of counting all syllables per 100 words. Such an approach was found to correlate .95 with the scores produced by the Flesch formula. Consequently, Farr, Jenkins, and Patterson suggested that their reading ease formula be substituted for Flesch's. Their formula is:

RE = 1.599 (number of one syllable words)
- 1.015 (mean sentence length in words) - 31.517

The Cloze Procedure

The cloze procedure, introduced by Taylor in 1953, was demonstrated to rank standard reading passages in the same order as did the Flesch and the Dale-Chall formulas. In the cloze procedure, readers are presented with samples of text, from which some words are deleted and replaced by blank spaces. The readers are requested to fill in the blank spaces with the words they think were deleted. To the extent that the author uses the words that the reader expects and understands, the reader will fill in the correct words. The technique assumes that readability is a direct function of the number of omitted words which the reader is able to fill in.

Taylor indicated that the ordering of cloze scores is maintained regardless of the system employed in word deletion. He used four different deletion systems on three standard passages, each of a different difficulty level, and found that the rankings were the same despite the deletion system employed. The four deletion systems involved deleting every fifth word, every seventh word, every tenth word, and 10 per cent of the words at random. Others have reported that deleting 20 per cent of a passage will yield sensitive measures.

The cloze method is free from many of the disadvantages of the traditional readability measures. It can be applied more appropriately to highly technical and unusual materials. Very technical material might be rated as difficult by the Flesch and Dale-Chall formulas, but not by the

cloze technique, if the subjects reading the passage were trained in the subject matter area. Conversely, the readability of the writings of authors such as Gertrude Stein, who write in short sentences with relatively simple vocabulary, but whose style is such that the material is difficult might not be accurately reflected by the Flesch and Dale-Chall techniques. The cloze test might be expected to reflect accurately the difficulty of such reading passages.

Taylor reported correlations of .70 and .80 between cloze scores and comprehension scores received by Air Force trainees on Air Force technical material. A study performed by Bormuth indicated that a 30 per cent cloze score was associated with a 75 per cent comprehension test score; 50 per cent cloze score was associated with 90 per cent comprehension score.

In a study performed at Applied Psychological Services, a series of norms was developed for use in evaluating the readability of submarine sonar manuals through the cloze technique. In order to obtain normative data, the researchers administered a representative sample of submarine sonar manuals, in accordance with the cloze procedure, to a representative sample of sonar technicians. The cloze scores achieved by the sonar technicians constitute the normative standard against which scores achieved by similar groups on other manuals may be compared. Thus, the norms allow an answer to the question, "Relative to the group of sonar technicians in the normative sample, what percentage of technicians can be anticipated to achieve a given comprehensibility score on this new manual?"

The inherent advantages of the cloze procedure are: (1) scoring ease, (2) scoring reliability, (3) ease of application to nonstandard material, and (4) accounting for the reader's interest in and prior knowledge of the content. The disadvantages of the procedure are: (1) cloze is a measure of readability, not a predictor of readability, (2) a sizeable sample of subjects is required, and (3) it may not reflect all types of comprehension.

The Dale and Chall Method

The readability formula developed by Dale and Chall was based on the average sentence length, and the "Dale score." The "Dale score" is the relative number of words in the passage that do not appear on a particular list of words, a list of 3,000 words known to 80 per cent of a sample of fourth graders. The Dale and Chall readability formula is:

$$X_{c50} = .1579 \text{ (Dale score)} + .0496 \text{ (sentence length)} + 3.5365$$

where:

X_{c50} = the reading grade level of an individual able to answer 50 per cent of the comprehension test questions correctly

Reading Level Determination

Several of the readability formulas discussed above predict readability in terms of reading grade level (RGL). RGL may be determined through several different approaches. The formulas above predict RGL as measured by standard tests of reading ability, for example, McCall and Crabbs' Standard Test Lessons in Reading. These tests may be used for groups or individuals. An annotated list of these tests is available in The Sixth Mental Measurements Yearbook (Buros, 1965).

Although the standardized tests of reading comprehension have demonstrated validity and reliability, there are more efficient techniques available for use in the military situation. Madden and Tupes provided a method for estimating RGL from Air Force aptitude test scores. They found that the general aptitude index (AI) of the Airman Qualifying Exam (AQE) correlated .70 with RGL. This high correlation was largely attributed to the reading vocabulary subtest in the AQE. In predicting the RGL of an individual Airman, one of the following three formulas may be employed, depending upon the Airman's career field:

Administrative: RGL = .0437 (Gen AI) + .0501 (Ad AI) + 5.0730

Mechanical: RGL = .0991 (Gen AI) + .0085 (Mech AI) + 5.0459

Electronic: RGL = .0743 (Gen AI) + .0222 (El AI) + 4.6088

Another regression equation for predicting an individual's RGL, in the military, was developed by Caylor, Sticht, Fox, and Ford (1972). Their formula uses the individual's Armed Forces Qualification Test (AFQT) score. The formula is:

RGL = .75 (AFQT score) + 5.52

Additional Reading List

Chall, J. S. Readability: An appraisal of research and application. Columbus: Bureau of Educational Research, Ohio State University, 1958.

Klare, G. R. The measurement of readability. Ames, Ia.: Iowa State University Press, 1963.

Osgood, C., Suci, G., & Tannenbaum, P. The measurement of meaning. Urbana: University of Illinois Press, 1957.

Pool, I., De Sola Trends in content analysis. Urbana: University of Illinois Press, 1959.

Szalay, T. G. Validation of the Coleman readability formulas. Psychological Reports, 1965, 17, 965-966.

Williams, A. R., Jr., Siegel, A. I., & Burkett, J. R. Readability of textual material--a survey of the literature. Lowry Air Force Base, Colorado, AFHRL-TR-74-29, in preparation.

Yoakum, G. A. Basal reading instruction. New York: McGraw-Hill, 1955.

II. APPLICATIONS FOR INCREASING READABILITY

Communication through writing, in a learning situation, is difficult because of the many complexities involved. Often the language that must be used is technical, students' reading skills are low, and the subject matter material is difficult to present in an interesting fashion. Despite the difficulties inherent in transferring information through the written word, textual presentation is a communication technique that trainers must use. Written material is considered maximally effective when a student can gain a clear understanding of the content after a single reading. Accordingly, it is imperative for the writer to use language that will be meaningful to the student—language that is easily understood.

This section suggests methods for increasing the readability of material prepared for military training purposes. Because multimodal presentation of the material to be learned has been recently suggested in a number of contexts, this concept will also be discussed. Although blanket statements about the benefits that can be gained from a multimodal presentation of material cannot be made for all types of material or for all types of learners, consideration should be given to multimodal presentations when: (1) the material lends itself to presentation in more than one mode, (2) the increased benefits outweigh the increased costs (section IV), and (3) reading grade levels of personnel to be trained are low.

Gains to be Anticipated

The gains that can accrue from an effective presentation of training material are:

- less time to achieve the required skills and knowledges
- less effort required to learn the material

These factors translate directly to increased effectiveness through lower costs both in the training center (through the savings in time to train the student) and in the field (through the diminished amount of time required for on-the-job training).

Time for Comprehension

There are certain types of material that are more easily grasped and retained when they are presented in writing. Some types of material require time to be absorbed. The time needed to absorb the material can be taken by the student if he reads the material, but he may not have the time he needs if he is unable to control the rate of presentation. For example, the student must pace himself according to the speed set by the instructor in the lecture, whereas he can move at his own pace when reading written material.

Logical and Precise Writing

Logical and precise writing is the key to improving readability. Quite obviously, readability will be improved if logic and precision are considered in the context of the structural factors suggested by the various readability formulas.

When preparing readable material, it is important to know the reading ability of those for whom the material is being prepared. Certainly, a writer of textual material would not write at an eigth grade reading level if he were preparing course material for college seniors. On the other hand, it would be folly to write a discourse at the sixteenth grade reading level when it is to be read by enlisted Airmen taking a technical training course. Therefore, any "rules" for preparing written material must be tempered by the background and reading grade levels of the readers.

Flesch suggests:

1. preparing the material in a personalized manner—increase the number of personal words (e.g., I, you, he, him, she, her; words of masculine and feminine gender such as John, Miss, mother, brother; and group words such as people) and increase the number of personal sentences (e.g., spoken sentences in which quotation marks are used; sentences addressed to the reader as in "do this" and "don't do that;" exclamations; and incomplete sentences)

2. discussing all points presented

3. shortening the length of words and sentences (if long technical words have to be used, make sure the meaning will be clear to the reader even if a definition must be included)

4. using punctuation as an aid to the reader

5. discussing points in chronological order or in order of increasing difficulty

6. avoiding excessive wordiness

7. providing anecdotes or illustrative examples wherever possible

8. saving the most important points for last; many learners remember best what is presented last

Learning from textual material will also increase when the following are included:

1. an introduction outlining the points that follow

2. a summary at the end

3. major and minor headings to guide the reader

4. transitional paragraphs which lead into subsequent topics

5. statements telling the reader what the important points are—telling him what to remember

In addition to those aspects outlined above, which all serve to improve clarity, the writer should present one point at a time and stick to that point. Deviations from the main theme must be avoided and important points within the theme must be emphasized. Emphasis can be accomplished in several different ways; e.g., underlining, italics, repetition, a statement such as "this is an important point," and devoting more space to the points which are to receive greater emphasis.

A conversational writing style also enhances readability because one tends to present his ideas simply when speaking, but may become a little more complex when committing his thoughts on paper. As an example, an instructor might say to his class "Tomorrow you will have a 10 minute multiple choice test on Chapter 2." The message is clear and concise. If a literary written instruction was given, it might be worded, "In order to determine your level of achievement on the subject of biological warfare, it is recommended that you be examined at the completion of the subject. Therefore, a paper and pencil test will be administered." The consequences of such an instruction could be sizeable to the instructor (if the students require clarification from him personally) or to the students (if they appeared at the next session of the class unprepared for the test). Although written instructions may be better than oral instructions because they are available for additional reference, they should be direct, concise, and follow the style of the spoken version in order to be effective.

Other suggestions which will assist the writer in preparing more effective written materials are:

1. Do not begin a sentence with the indefinite pronoun "this." "This" refers to something in the preceding sentence or paragraph, but the specific antecedent may not be clear to the reader. Since the goal is to increase comprehensibility, repeat a word or concept rather than chance being unclear to the reader.

2. Avoid using colloquial expression which may be meaningful to certain groups of people but not to others; e.g., "groovy," instead of "very acceptable."

3. Use parentheses to set off explanations. Do not enclose one set of parentheses by another; use brackets to enclose parentheses; e.g., 10[(a+b)(c+d)].

4. Use abbreviations only after a full explanation of the abbreviation has been provided, except where the abbreviation is very widely known; e.g., 12 a.m. An abbreviation is often defined in parentheses after the first use. The abbreviation may be used alone thereafter.

5. Use only those symbols that are very common; otherwise spell out the symbol. Examples of some common symbols are: $, ¢, #, =, etc. If less popular symbols are needed, define them for the reader.

6. Do not omit an article for the sake of brevity; e.g., "The students and the instructors were invited to the party," not "The students and instructors were invited to the party."

7. Do not omit the word that; e.g., "Lt. Smith felt that he should speak to Pvt. Jones," not "Lt. Smith felt he should speak to Pvt. Jones."

8. Do not omit a verb; e.g., "He finished the course on June 6 and was tested on June 8," not "He finished the course on June 6 and tested on June 8."

9. Place modifiers as close to the word they modify as possible; e.g., "He studied diligently for three hours," not "He studied for three hours diligently."

10. Avoid splitting infinitives unless the sentence would be awkward; e.g., "She wanted to return promptly after Christmas," not "She wanted to promptly return after Christmas."

11. Avoid using the word and to connect two unrelated ideas; either rephrase the sentence using other connectives or write two sentences. For example, "Since the instructor taught the material, the students were able to answer the question," or, "The instructor taught the material. Thus, the students were able to answer the question;" but not "The instructor taught the material and the students were able to answer the question."

12. Use the active voice rather than the passive voice. For example, "Use a screwdriver to help you raise the lid;" not "A screwdriver should be used to help raise the lid."

Word and Sentence Simplification

The selection of words is crucial to preparing readable copy. A general principle to follow is to select words that have a high frequency of occurrence and are thereby familiar to the readers. In addition to selecting familiar words, the words used should be:

a. short
b. nontechnical (where possible)
c. American sounding rather than foreign sounding
d. concrete rather than abstract

The next logical consideration is sentence structure. Here, the basic factor is the length of the sentences. The characteristics of the reader will be the largest single factor in determining how long or short sentences should be. Generally, shorter sentences are more readable. However, other sentence structure characteristics should be considered in making written material more readable. The characteristics to be avoided are:

1. prepositional phrases (e.g., The two roads of which I speak are...)

2. compound sentences (e.g., He campaigned for nomination, ran for election, and won the vote.)

3. complex sentences (e.g., I would have passed the test if I studied more.)

Since the more complicated sentences are usually longer than simple sentences, it can be demonstrated that sentence length is directly related to sentence structure. Considering that sentence length is easier to analyze than sentence structure, length is the more frequently used factor in readability measurement.

The following examples represent excerpts from Air Force manuals, as they appeared in their original form and then in a modified version. The modified versions demonstrate the effects word and sentence simplification.

Original: The Air Force uses two basic methods—prepost and post-post—for issuing supplies. A prepost issue is processed through the computer before the property is actually selected from storage. A post-post issue made before the issue request is processed through or into the computer. A simple way to determine which method is being used is that prepost documents are machine printed and post-post documents are normally handscribed.

Modified: The Air Force uses two methods for issuing supplies. These are prepost and post-post. A prepost issue is processed by the computer before the supplies are taken from storage. A post-post issue is processed by the computer after the supplies are taken from storage. You can easily tell which method is being used because prepost order documents are printed by machine. Post-post order documents are usually written by hand.

Original: Post-post issues from the base service store are accomplished in the same manner as normal warehouse post-post issues. The AF Form 2005 and the DD Form 1348-1 are handscribed. Unlike the single item prepost issue, the property is not removed from stock until the document arrives at the BSS. If the entire quantity requested is not available, the quantity actually issued is entered in the quantity block of the DD Form 1348-1. The property and documents are forwarded to the pickup and delivery section for delivery and document distribution.

Modified: Post-post issues from the base service store are made the same way as regular warehouse post-post issues. The AF Form 2005 and the DD Form 1348-1 are filled in. The procedure is different than that used for a single item prepost issue. Supplies are not removed from stock until the issue form arrives at the BSS. If you don't have enough of an item in stock to fill an order completely, write the amount that you can supply in the quantity block of DD Form 1348-1. The items and forms are sent to the pickup and delivery section. These people will deliver the items and distribute the forms.

Original: Copy 1 of the AF Form 1991 for a location addition is forwarded to the warehouse remote operator for input, or to the punch card accounting machine (PCAM) unit for keypunching and input via the card reader. Copy 2 is held in a suspense file, pending receipt of the output notice or its appearance on the daily document register. If the input is by a remote device, an output notice is produced on the remote. If the input is made by the card reader, no printed notice is produced.

Modified: If the new warehouse location input to the computer will be through remote, send copy 1 of the AF Form 1991 to the warehouse remote operator. An output notice will be produced on the remote. If the input will be through a card reader, send the form to the punch card accounting machine (PCAM) operator. He will keypunch and enter the information. No output notice will be produced in this case. Keep copy 2 of the AF Form 1991 in a suspense file. When you receive the output notice, remove it from the file. You also remove copy 2 from the suspense file if that location input appears on the daily document register.

Repetition-Redundancy

Repetition is an important concept in readability-comprehensibility. Repetition reduces uncertainty in the reader by helping him to remember important points.

To increase redundancy, it is not necessary to repeat specific points verbatim or even to repeat the same point using different words. In some cases this could lead to monotony; those who wish to reread a specific section could do so. Increasing redundancy can be accomplished by writing about a specific issue from another point of view. Redundancy can also be increased by using words in a sequence; e.g., he is loud, dumb, and vulgar.

Textual Supplementation

Reading is a complex process involving the reader's education, reading techniques, intellect, and emotional status, as well as the characteristics of the textual material itself. Other factors affecting the ability of the reader to read quickly and understand textual material are the reader's visual acuity, his physical condition, his educational background, cultural factors, distractions present, the reader's ability to conceptualize language, to think symbolically, and to organize the material meaningfully. Diagnosis of a reading disability requires specially trained persons. A student's inability to master the written word may be overcome, to some degree, by supplementing the textual material with auditory tapes, pictures, graphs, and the like.

Auditory Supplementation

Although listening has an advantage over reading because it contains intonational nuances, individual differences among listeners exist insofar as rate of listening comprehension is concerned. Listening to tape recordings is somewhat like listening to a lecture, except for the formality of the classroom environment. However, the student can stop the tape at any time to think over a point. And, he can return the tape at an earlier point in case he has missed the point. Most research indicates that no student will do worse with taped material than with written texts and that some students will do better with taped materials as compared to written texts.

In preparing the material for a recording, the problem of making the material comprehensible to the learner is as important as for written material. The subject matter must be efficiently organized if the material is to be meaningful to the listener. Learning from tape recordings will be enhanced by the following:

1. limit the amount of information on any one tape
2. repeat key points at strategic places in the recording
3. keep the material at the level of comprehension of the students
4. permit discussions and questions before and after the student listens to the recording
5. prepare the student for the recording by telling him of the subject matter to be covered, its purposes, and how he will be able to use the information
6. save the most important points for presentation near the end
7. compliment the students by leaving them with a note of encouragement

The speaker on a tape recording should be skilled in the techniques of verbal delivery; i.e., someone trained in the area. Since the speaker on a recording cannot take advantage of his body and other audio-visual aids, he will have to concentrate on different types of verbal techniques to project his ideas and purposes. Some of the key points in making a good delivery are:

1. present the material in a lively and animated fashion
2. convey a sense of enthusiasm and belief in the importance of the subject matter. The material must be learned and understood by the speaker in order to deliver the talk with conviction.
3. use a pleasant tone and vary the pitch and force. The rate should be carefully controlled as should the degree of loudness.

4. use pitch, tone, and rate variation for emphasis

5. use inflections--a downward inflection at the end of a complete thought (as is typically used at the end of sentences); an upward inflection at the end of a question (often associated with surprise, doubt, uncertainty). Do not use rising inflections at the end of complete thought units, as it presents a sing-song quality

6. use short, crisp sentences

7. use a voice that is free from any unusual quality (e.g., harshness, nasality, breathiness, hoarseness, throatiness)

Illustration Supplementation

There are other alternative techniques available for presenting information that one might normally present in prose. These alternative approaches may result in improved comprehension and/or performance. Although an instructor may use what he considers to be simple and concise language (and the writer or reading material does the same) communication might be less than complete without visual presentation. Visual presentations show the learner exactly what the writer is talking about or what is being referred to in the written material. Illustrations are held to be essentially demonstrations but in a remote fashion. They save time because fewer words have to be written to describe a situation and, accordingly, fewer words have to be read.

The following should be taken into consideration in preparing illustrations:

1. avoid unnecessary artwork since it may serve only to distract

2. colors, if used, should be in sharp contrast with each other

3. if temporal sequencing is involved (e.g., as in a step-by-step procedure) and only that part of the equipment is to be visible which applies to a particular step, then overlays on transparancies should be used

4. limit the number of callouts or identifiers

5. connect identifiers to the correct point in the illustration with a leader line

6. limit the callouts to those referenced to in the text

7. use directional arrows to help the reader orient himself with respect to the illustration

8. all text appearing on the illustration should be in an area which will not affect the readability of the illustration, its callouts, or identifiers

9. simplify as much as possible (an overcrowded illustration is confusing)

10. use size and placement for emphasis

11. discuss the important aspects of each illustration in the text and state how to apply the information

12. keep the illustration interesting--from the reader's viewpoint

13. choose a simple, easy to read letter style

Preparing effective illustrations, a field in itself, requires professionals and experts. Commercial and industrial artists, engineers, and draftsmen are often consulted in these matters. However, since these professionals may not be subject matter experts, effective illustrations will represent the merging of instructors, who know the subject matter and message to be delivered by the illustration, with professionals who have the artistic talent.

Flowcharts and Tables

Flowcharts and schematics also represent forms of textual supplementation. The flowchart is one form of illustration which is helpful when several complex, serialized points have to be made or when the subject matter pertains to interactive sequences of acts that should be performed under different conditions. The flowchart is more easily understood, more accurately used, and more quickly interpreted than its prose counterpart. Figure 1 presents a set of instructions in prose form.

Instructions in Prose

Place switch S-1 on the tester in the "on" position. Probe connections are required from scope input A to TP3 on the LLM. Adjust the OUTPUT LEVEL on the LLM so that the amplitude is 5 v. P/P. If there is no output at TP3, check for a signal at V101, pin 1. If there is a signal at V101, pin 1, then check for a signal at V102, pin 2. If there is no signal at V102, pin 2, there is trouble in the filter network. There is trouble in either V102, T101, or associated circuitry if there is a signal at V102, pin 2. If there is no signal at V101, pin 1, check for a signal at V101, pin 2. If there is a signal at pin 2, the trouble is in the ..., but if there is no signal at pin 2, then check V103, pin 3, for a signal

Figure 1. Sample of instructions in prose.

These same instructions are also presented as Figure 2 in the form of a flowchart. It can be seen, from Figure 2, that the flowchart will help the reader who is searching for a specific piece of information. It also provides an integrated picture so that relationships and orderings among the parts can be readily perceived.

BEST COPY AVAILABLE

STEP 1

SWITCH S-1 ON TESTER	PROBE CONNECTIONS		WAVEFORM
	FROM SCOPE	TO LLM	
CH-1	INPUT A	TP3	
NOTE: ADJUST OUTPUT LEVEL CONTROL ON LLM SO THAT AMPLITUDE IS 5V. P/P			a = 5V b = 513 ± 26 μs

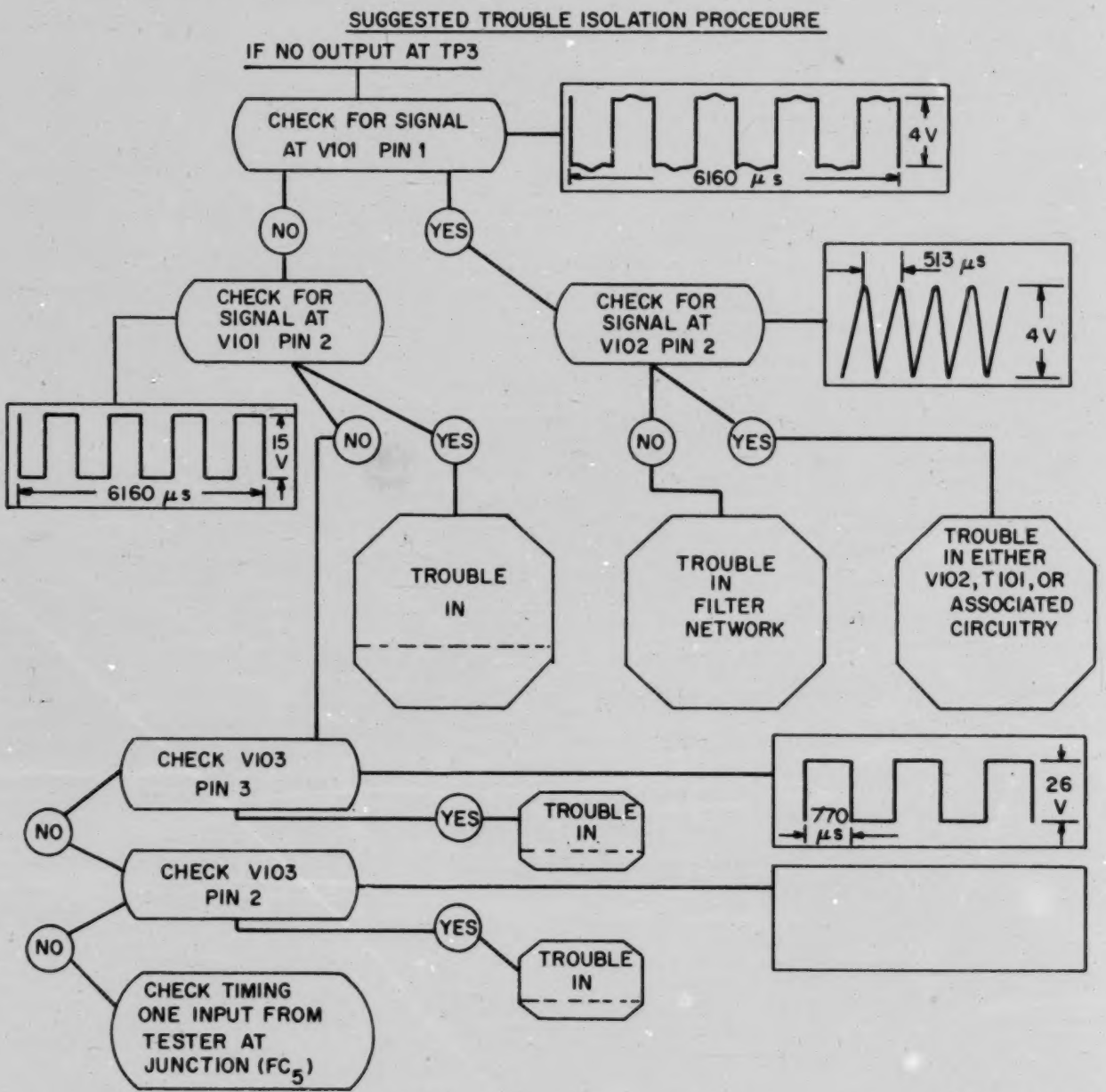


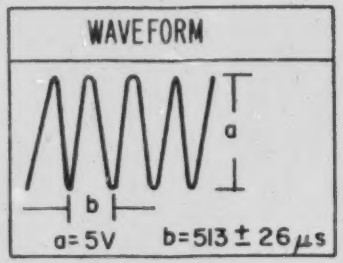
Figure 2. Example of flowchart to represent instructions of Figure 1.

Another alternate to prose is tabular presentation. Figure 3 is a two-dimensional table, presenting the same information as in Figures 1 and 2. While complex information is difficult to present, regardless of the format, it is evident that for the information presented in Figures 1, 2, and 3, the prose version is the least comprehensible.

The advantages of presenting quantitative data in a tabular format are that tables are: (1) economical, (2) systematic, (3) concise, and (4) comparisons are facilitated within and among tables. The following will help in preparing comprehensible tables:

1. title the table with a brief descriptive statement of the contents
2. legend each column
3. legend each row
4. separate classifications (i.e., between rows and between columns) by ruling in lines or leaving wide spaces
5. identify the units that are used for the numbers in the column (e.g., dollars in thousands, etc.)
6. use footnotes to point out any unusual circumstances or information that cannot be placed in the body of the table

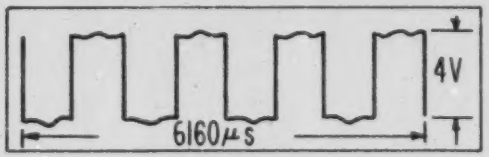
LLM Test Procedure



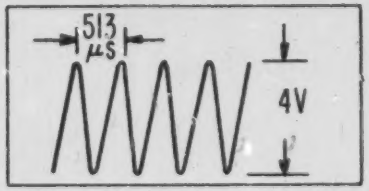
1. Tester S-1 to ON
2. Probe connections from scope input A to LLM, TP3
3. If no output at TP3, then

Step	Signal-No Signal	Action	Indication
1. Check V101, pin 1	Signal	Go to step 2	
	No signal	Go to step 3	
2. Check V102, pin 2	Signal		Trouble in V102, T101, or associated circuitry
	No signal		Trouble in filter network
3. Check V101, pin 2	Signal		Trouble in ---
	No signal	Go to step 4	
4. Check V103, pin 3	Signal		Trouble in ---
	No signal	Go to step 5	
5. Check V103, pin 2	Signal		Trouble in ---
	No signal		Check TIMING ONE input from tester at junction (FC_5)

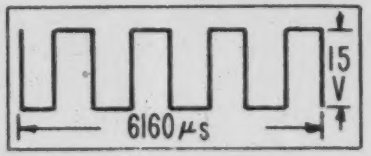
Step 1



Step 2



Step 3



Step 4

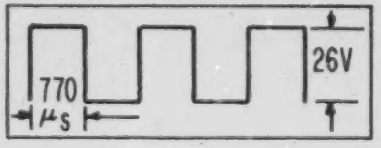


Figure 3. Example of tabular presentation of instructions.

BEST COPY AVAILABLE

Graphs

Graphs are of particular value when the writer wishes to display the general shape of a function, or where trends or relationships between quantities are being described and contrasted. There are several different forms of graphs. The three most popular are the line graph, the bar graph or histogram, and the circle or pie graph. The graph should be placed close to the place where it is mentioned in the text.

In the typical line graph, when the functions portrayed are continuous, a set of values is described on the horizontal or x axis and another set of values is described on the vertical or y axis. When plotting the results of an experiment, the independent variable, by convention, is laid off on the horizontal axis and the dependent variable is laid off on the vertical axis. The following will help in the preparation of comprehensible line graphs:

1. draw the horizontal and vertical axes with heavy and distinct lines

2. label the axes in terms of the units used (e.g., dollars in thousands, percentage, etc.)

3. title the graphs descriptively but as concisely as possible

4. place the title below the graph; never place it within the grid or plotting area

5. select convenient scales for both axes; the scale selected must remain the same from the beginning of the axis to the end

6. select units that are easy to divide (e.g., 1, 2, 5, 10, 50, 100, etc.) so that intermediate points will be relatively easy to read

7. construct the axes so that the vertical axis is 2/3 to 3/4 as long as the horizontal axis

8. arrange the graph so that it is read from left to right and from the bottom to the top

9. assign the value of zero to the point where the two axes meet; if the data are such that there is a large gap between zero and the next value, the axis is discontinued between zero and the first value by using the symbol ≈ (see Figure 4)

10. use dots or crosses to plot values corresponding to the intercept points on each axis

11. if several curves of different importance are plotted in the same graph, make the principal curve the most prominent (e.g., use a thicker line)

12. limit the number of functions plotted (curves on the graph) to four; try to avoid a cluttered appearance

13. provide legends wherever an explanation or an identification of different lines is required

An example of a line graph is presented in Figure 4.

The second most commonly used type of graph is the bar graph. The bar graph is a representation of discrete numerical quantities and is useful for displaying differences among these numerical quantities. For example, if the frequency of students from various states were to be plotted on a bar graph, the states would be listed on one axis and frequencies or percentages on the other. There are no rules regarding the axis to use for the different variables, or the lengths or heights of the bars. If there is no logical or ordinal relationship among the discrete entities plotted, then they may be arranged from lowest to highest. An example of a bar graph is displayed in Figure 5. The graph should be titled. By convention, the title is placed below the graph. Provide supplementary descriptive information that is required for understanding the graph in a key or legend.

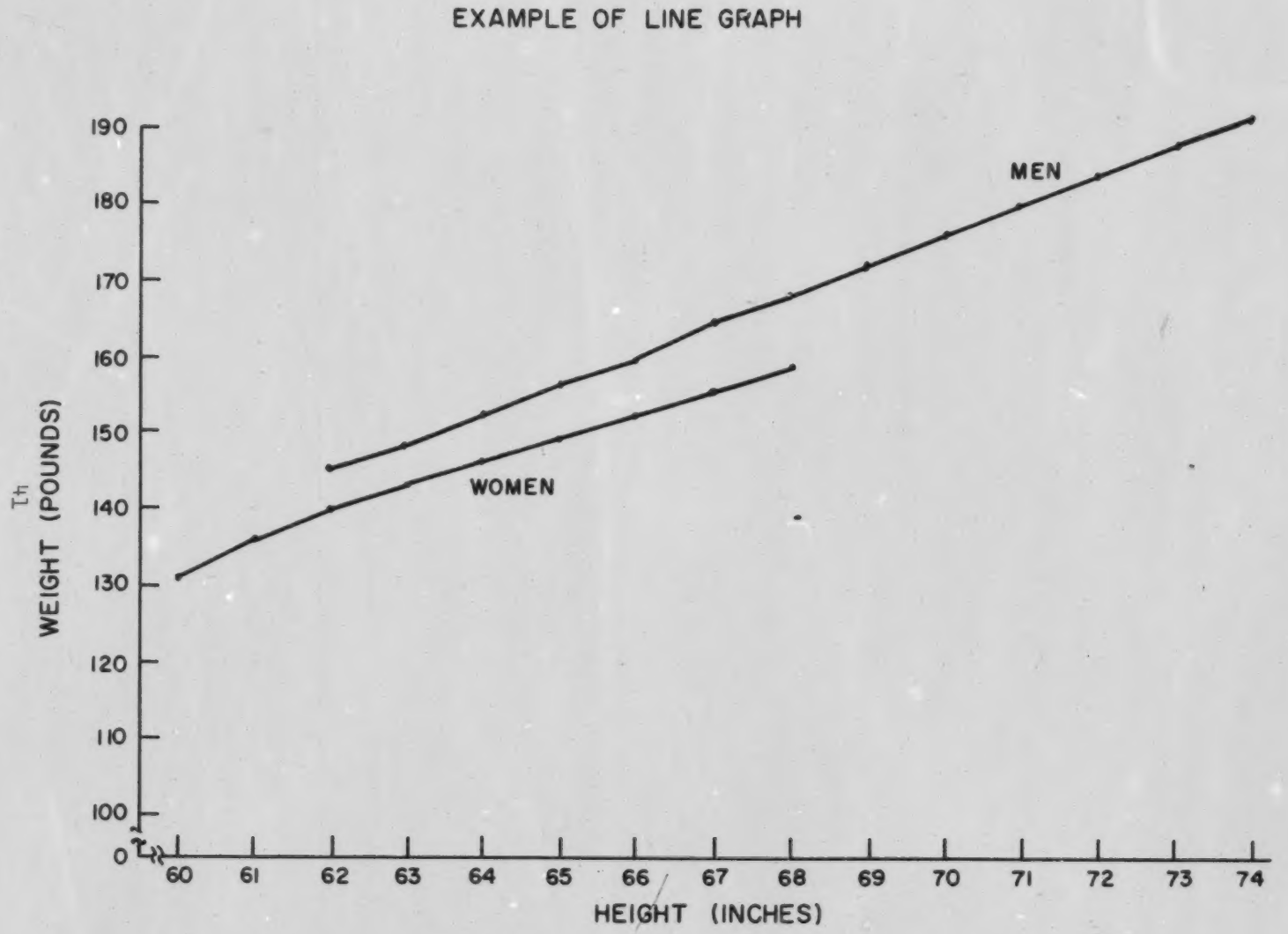


Figure 4. Weight as a function of height for men and women between the ages of 18 and 64.

BEST COPY AVAILABLE

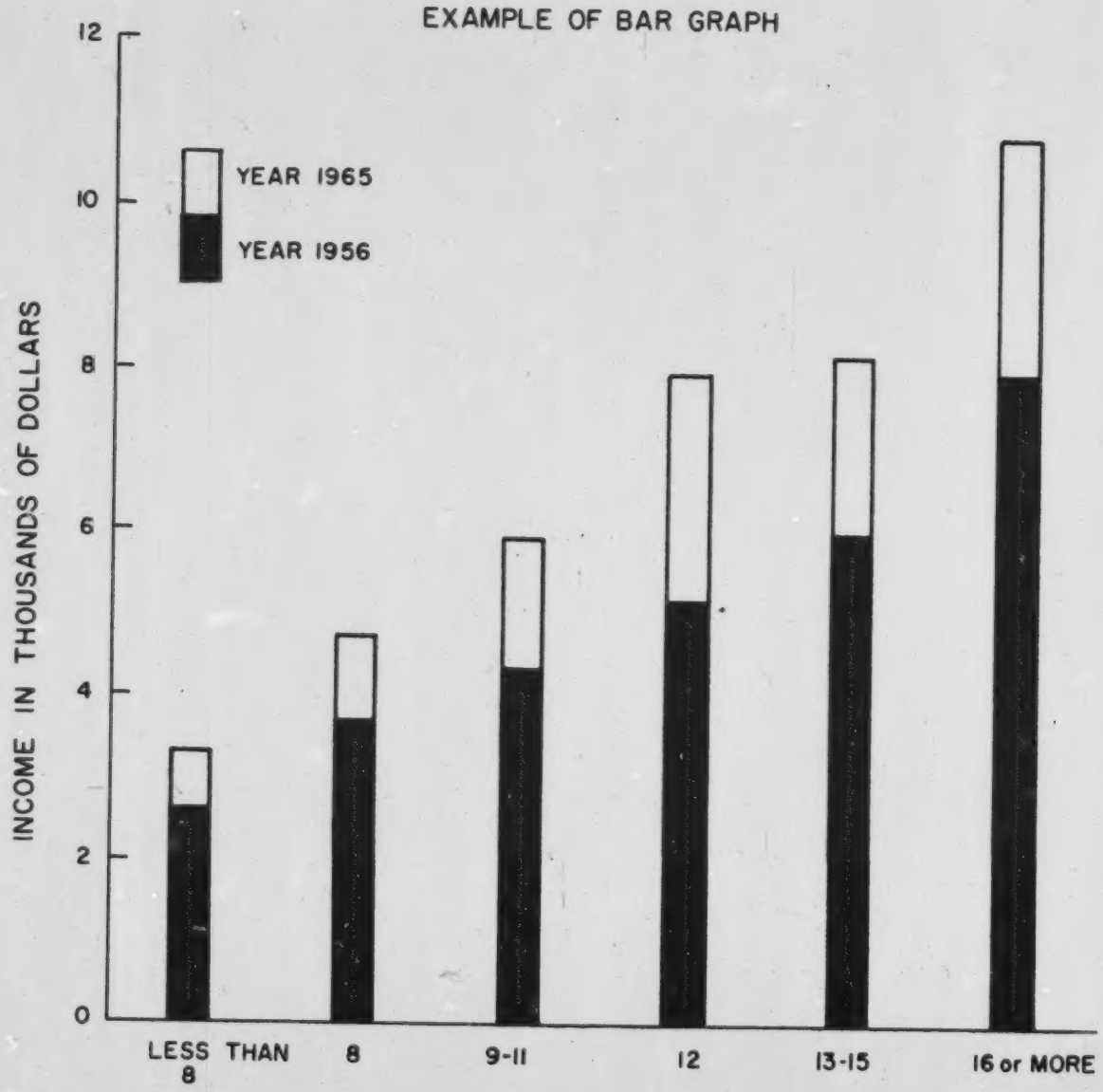


Figure 5. Annual mean income of males (25 years and older) in 1956 and 1965 as a function of years of education.

BEST COPY AVAILABLE

The circle or pie graph displays differences among the data in terms of different size sectors or segments. It is useful for demonstrating the relationship among parts of a whole. It is a simple graph to draw, and easy to understand. To prepare an effective circle or pie graph:

1. title the graph—below the circle
2. legend each sector
3. include in each sector the proportion of the total circle that is represented by the sector
4. make the size of each sector directly proportional to its proportion of the whole circle
5. use a protractor for drawing the circle and for plotting the percentages of each segment

An example of a circle graph is presented in Figure 6.

Schematic Diagrams

Schematics have been broadly used for equipment maintenance instruction and on-the-job maintenance support in both military and commercial settings. The IMM (Integrated Maintenance Manual) and the SIMM (Symbolic Integrated Maintenance Manual) by the Navy and Coast Guard are examples of military manuals that have incorporated the use of concise, consistent, and well-laid out schematics of equipment. PIMO (Presentation of Information for Maintenance and Operation) is the Air Force version for presenting technical order type instructions and information. The information in PIMO emphasizes how to do something, rather than what should be done.

BEST COPY AVAILABLE

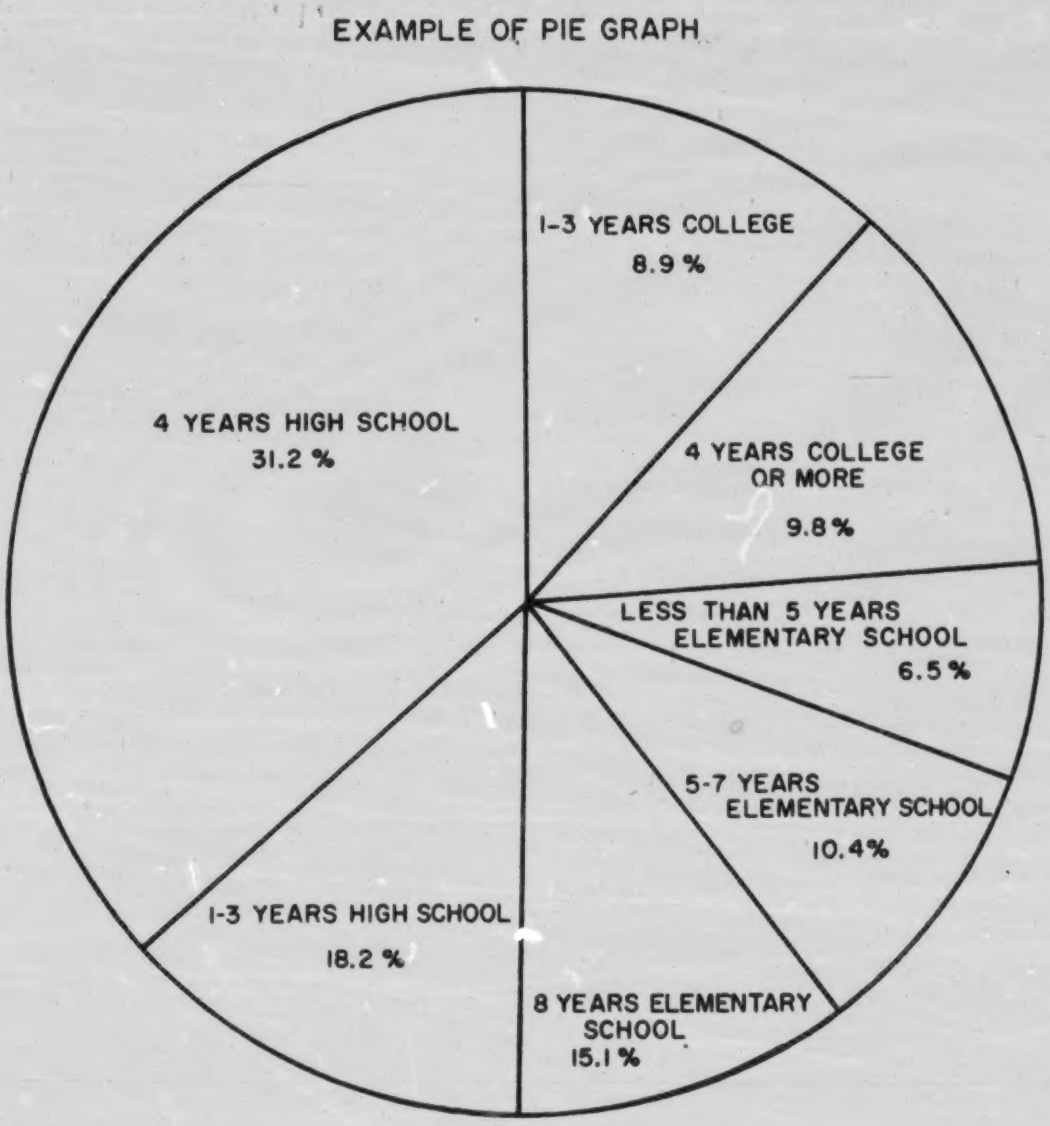


Figure 6. Percentage of United States' population completing various years of schooling, 1966.

Comprehensible schematics will include:

1. easy to understand layouts
2. clearly demarked circuits and components
3. maintenance data
4. the same functional divisions as the equipment
5. aid for locating components (e.g., grid underlay)
6. cross-sectional schematic views of mechanical devices
7. coding (e.g., flow arrows to show direction of flow)
8. wiring and cabling interconnections

When text is included with the schematic, it should be:

1. placed on the page facing the schematic
2. readily associated with the schematic portion to which it pertains
3. short but complete

Figure 7 presents a schematic that incorporates many of the principles outlined above.

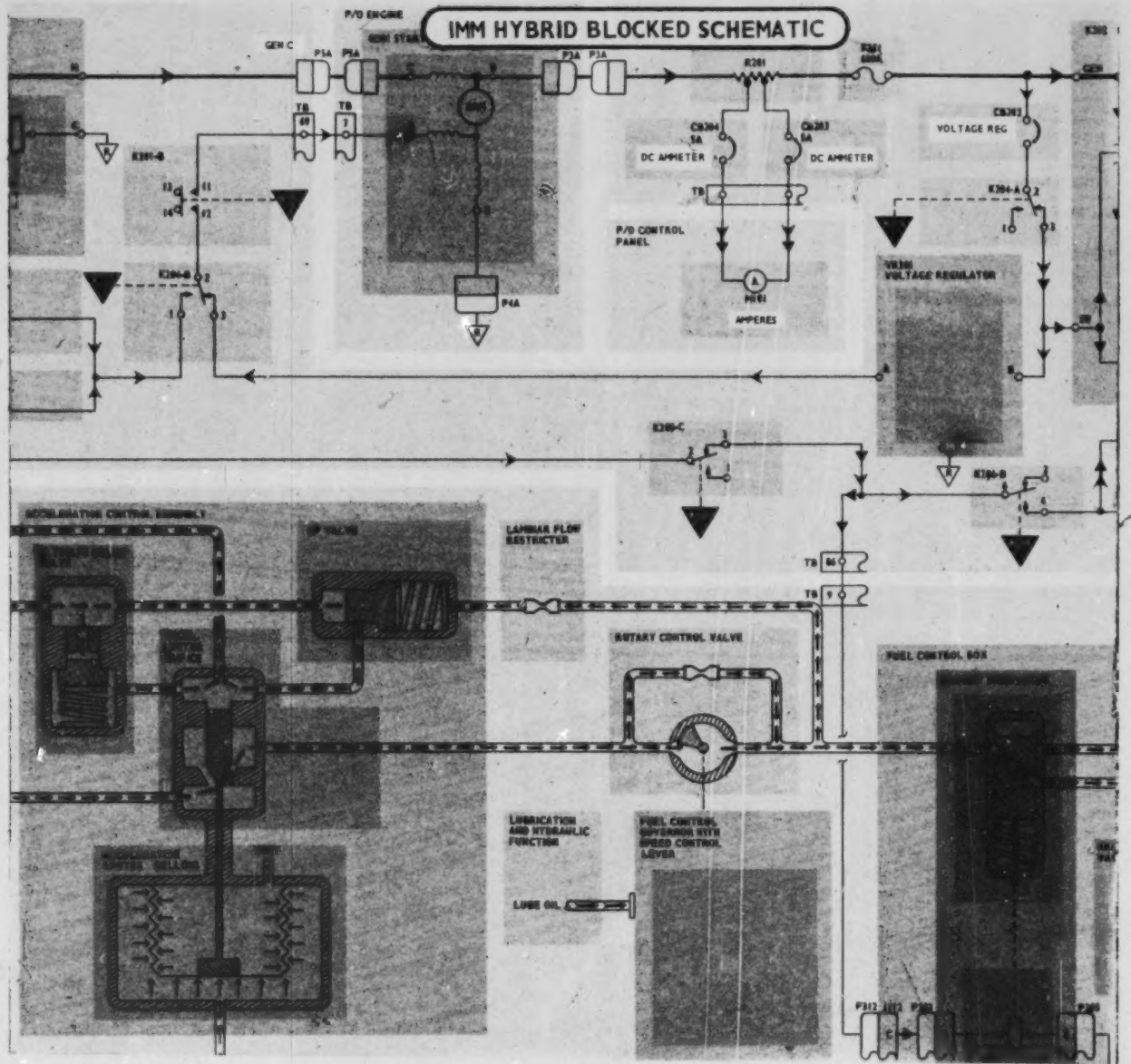


Figure 7. Example of schematic incorporating comprehensibility principles.

This excerpt presents detailed schematic level coverage of a hybrid equipment consisting of electrical and fluid control components.

Features include:

- PRECISE HARDWARE DEFINITION
  Hardware boundaries are defined by gray shading.
- PRECISE FUNCTIONAL DEFINITION
  Functional elements (circuits and mechanical components) are defined by blue shading.
- INTEGRATION OF ELECTRICAL AND MECHANICAL DEVICES
  Electrical, mechanical, hydraulic, and pneumatic information is integrated to increase understanding.
- IMPROVED PRESENTATION OF MECHANICAL DEVICES
  Valves and other mechanical devices are presented using cross-sectional schematic views to enhance understanding.
- IMPROVED SIGNAL CODING
  Coded flow arrows show content of fluid flow lines. Coded arrow-heads denote type of electrical signals.
- IMPROVED TECHNICIAN INTERFACE
  Front panel controls and indications are denoted by a cutaway white background.

In addition, this diagram has a facing page block text diagram to provide complete description of operation.

The use of cross-sectional schematic views is also applicable to equipments using microwave devices.

The total hybrid diagram represents a complete functional entity, Thus, the system is learned in an integrated manner enabling the technician to easily see the relationships between the different components of a hybrid system.

BEST COPY AVAILABLE

Additional Reading List

A manual of style. Chicago: The University of Chicago Press, 1964.

Anderson, U. A. Training the speaking voice. New York: Oxford University Press, 1942.

Bruning, J., & Kintz, B. Computational handbook of statistics. Glenview, Ill.: Scott, Foresman, 1968.

deKieffer, R. E., & Cochran, L. W. Manual of audio-visual techniques. New York: Prentice Hall, 1962.

Gunning, R. The technique of clear writing. New York: McGraw-Hill, 1952.

Haas, K. B., & Packer, H. Q. Preparation and use of audio-visual aids. 1955.

Prochnow, H. V., & Prochnow, H. V., Jr. The public speaker's treasure chest. New York: Harper & Row, 1964.

Tate, Merle W. Statistics in education. New York: Macmillan, 1964.

Thomas, P. E. L. A guide for authors. Springfield, Ill.: Charles C Thomas, 1961.

Wilmut, H., Chubb, G., Tabachnick, B. PIMD data preparation guidelines. Norton Air Force Base, Calif., AFSC TR 69-155, May, 1969.

Wright, P. Writing to be understood: Why use sentences? Applied Ergonomics, 1971, 2, 207-209.

III. "NONANALYTIC" METHODS FOR EVALUATING TEXTUAL METHODS

While use of the word counting (e. g., the Flesch or Dall-Chall technique) or associated (e. g., cloze) methods will be useful in evaluating the readability of textual materials, there are a number of nonanalytic techniques which will also be helpful to the material developer who wishes to evaluate his product. These nonanalytic methods are performance oriented. They consider the actual performance of the person who reads the material rather than the characteristics of the text itself. Generally, the nonanalytic techniques involve administering the written materials to a group and, then, testing the group to determine how much they have learned, or how much they comprehended about the subject matter from reading the materials.

It might be taken for granted that modifying the textual materials to produce a lower (and presumably easier) reading level will necessarily produce learning benefits. But it is also logical to assume that a two pound weight will fall twice as fast as a one pound weight. In the final analysis, some test of the effects of a materials modification on actual learning may be desirable. This chapter presents concepts and considerations important to such an endeavor. Such concepts and considerations include the design and conduct of such tests as well as the characteristics of the actual measurement instruments to be involved in the evaluation.

Design of Materials Evaluations

The evaluation of textual materials through nonanalytic methods will rely on the learning they produce, as measured by tests of the performance of the reader. Such tests may be administered immediately after the materials are read or after a time interval has elapsed. In the latter case, the final result will be confounded by the intervening activity of the reader. Thus, the intervening activity must be carefully controlled, or it will not be known whether any actual learning can be attributed to the materials themselves.

Because the design and analysis of a nonanalytic evaluation is a complex process, it is best to consult experts in research methods and statistical analysis before conducting such an evaluation. To perform a nonanalytic test, study guidelines, which are specific and not

subject to personal interpretation, should be developed in advance. The guidelines must describe fully the methods, procedures, data collection instruments, and statistical data treatment methods to be employed. Estimates of the time needed to conduct each aspect of the study should also be included. Example methods for conducting such evaluations are described below.

Single Group Methods

Assume that the textual materials for a course have been modified and that the materials developer wishes to know whether or not the modification produces any significant change in the amount of student learning. Assume that there are four blocks in the course involved. The original materials (O) could be administered to students when they are in blocks 1 and 3, while the modified (M) materials could be administered to the same students when they are in blocks 2 and 4. The class scores on end of block 1 and 3 tests (O materials) could then be compared with their scores on block 2 and 4 tests (M materials). One would anticipate higher scores on the M materials than on the O materials, and the difference between the two could be used as an index of the gain achieved by the modification.

In this type of study, each reader is exposed to each type of material. This avoids certain problems associated with control group studies. However, note that this type of design assumes that learning block 1 does not affect learning in block 2, etc., and that blocks 1 and 3 are at the same difficulty level as blocks 2 and 4. Such assumptions may not always be appropriate, and when they are not, this method should be avoided.

Control Group Studies

When the assumptions of the single group study cannot be met, control group studies can be employed. In the prior example, the students could have been split into two groups. In making such a split, it is essential that the two groups be equated on all variables which could affect the evaluative outcome. In the present context, reading grade level and background in the field of study might represent the critical variables on which the two groups would be equated.

There are two methods for establishing equivalent groups. One is to select the persons for assignment to each group from the total pool of available persons at random. For example, if an alphabetical roster is available, every other person might be assigned to one of the groups and the remainder would be assigned to the other group. The second method is to match, for the variables thought important for the study, as closely as possible the persons assigned to each group. Thus, if there is available a pool of 20 persons of known and varying reading grade levels and they are to be assigned to equivalent groups, the final matching might look as follows:

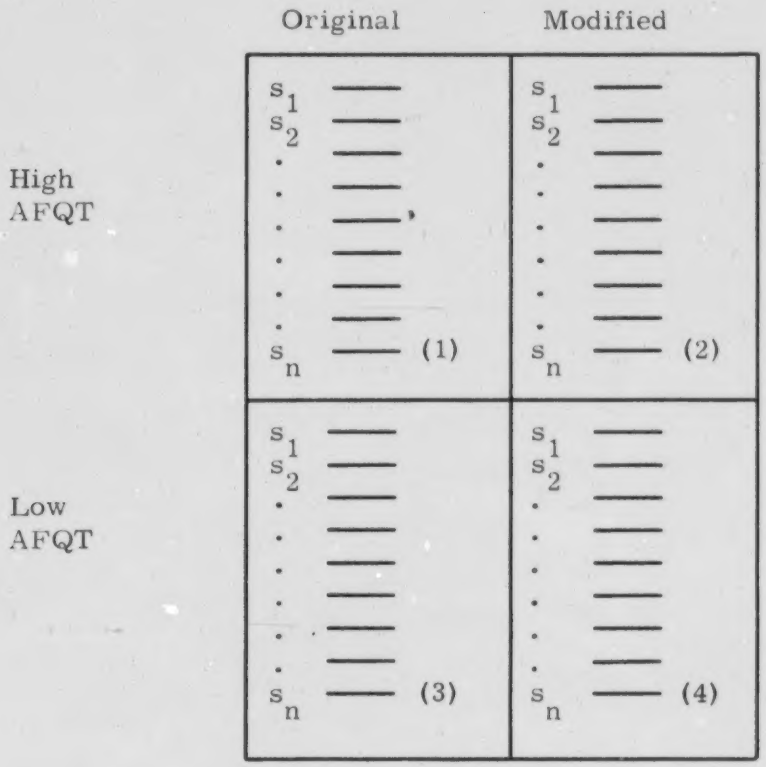
	Reading Grade Level of Matched Subjects	
Matched Pair	Group A	Group B
1	12.1	12.1
2	11.9	11.8
3	10.5	10.6
4	10.0	10.0
5	9.7	9.8
6	9.8	9.7
7	8.9	8.9
8	8.8	8.7
9	8.5	8.6
10	8.0	8.0

Having matched the groups, the control group technique involves administering the O materials (for all four blocks) to one group and the M materials (for all four blocks) to the other group and administering the appropriate tests. Possibly, such tests would be administered at the end of each block as well as at the end of all four blocks. Any differences between the test scores of the two groups (either on the individual block tests or the final examination) would then be attributable to the textual materials themselves.

Factorial Design

A still more powerful nonanalytic evaluative approach is through the factorial design. Using the same example, assume that the evaluator wishes to know whether or not there is a difference in the learning produced by the two types of materials and also whether or not this difference holds regardless of the general intelligence (AFQT score) of the student. Factorial designs permit the evaluator to simultaneously look

at the effects of two or more factors on learning. In the following example, AFQT is considered to be a factor, while difficulty of reading materials is considered to be a second factor.



In this example, the total group is split into two groups: high AFQT scorers and low AFQT scorers. Half of the high AFQT scorers receive the O materials and half receive the M materials. The same holds for the low AFQT scorers--half receive the O materials and half receive the M materials. The analysis (called analysis of variance) of the test scores of the four groups will tell the evaluator: (1) whether or not a difference in learning is produced by the two versions of the text, (2) whether or not high AFQT scorers make higher test scores than low AFQT scorers, and, most importantly, (3) whether or not there is a differential effect on the benefit to be gained by each type of material as a function of AFQT score.

Within single group, control group, and factorially designed evaluations, there are a large number of sub approaches. Accordingly, an expert should probably be consulted during the formulation of nonanalytic studies to insure proper experimental design.

Conduct of Nonanalytic Studies

No matter how powerful the design of the nonanalytic study, usable data will not be obtained unless most, if not all, of the following adequacy conditions are satisfied:

- control--all conditions in the situation and within the subjects which might affect the evaluation must be controlled; i.e., eliminated from the evaluation conditions or distributed equally over all conditions
- sample--the readers in the experiment should be fully representative of those who will eventually use the materials and a sufficient number of subjects must be included to allow stability in the resultant descriptive and inferential statistics employed
- question realism--the question(s) asked by the study must be drawn from sound concepts or theory
- assumptions--the assumptions must be clearly stated and acceptable
- objectives--the study objectives must be unambiguous and clearly defined
- personnel--the personnel who perform the study must possess sufficient training, experience, and sophistication to properly conduct the experiment
- stimuli--the reading materials included (stimulus representations) must be meaningful to or mirror the problem on hand, and the stimuli and stimulus combinations must cover an appropriate range of materials
- criterion weighting--if subscores are combined they must be appropriately weighted
- alternate methods--backup or alternate methods should be available, in case the favored approach must be abandoned
- time and costs--the time schedule must be realistic and compatible with available manpower; the cost estimate must be compatible with manpower, facility, and equipment requirements; contingency time and costs must be allowed

Characteristics of Useful Measure of Performance

Aside from concise and specific methods and procedures, readability evaluations should employ performance measuring instruments which possess sound measurement qualities. Recommendations from an evaluation can be no better than the data on which the recommendations are based. Decisions have to be made on the basis of the readability evaluative data. In order for a data set to be scientifically defensible, the measures on which the data set are based should meet most, if not all, of the following criteria:

1. Reliability--A measure is reliable to the extent that it will yield the same score when a person is measured with it on two separate occasions. If a measure does not possess this attribute, it is worthless for evaluation purposes. There are several methods for assessing the reliability of an evaluative measure and all of them involve the calculation of a correlation coefficient. Any elementary statistics textbook will instruct the reader in methods for calculating and interpreting a correlation coefficient.

2. Validity--Validity refers to the extent that a measure assesses what it intends or purports to measure. It is possible for a measure to be highly reliable, yet have little or no validity.

3. Comprehensiveness--Any useful performance evaluative measure must sample from the total range of reading material and not from a part of the range.

4. Objectivity--Only those aspects which can be objectively evaluated should be included in the measurement instruments. If the evaluator's personal biases can enter into the data, the measurements are subjective rather than objective. A measure is objective to the extent that different persons who use the same measure will give a student the same score.

5. Differentiation--Evaluation measures must be able to reflect differences in the variable(s) being measured. This allows decisions and comparisons to be made on the basis of known amounts by which one presentation of a written selection differs from another presentation of the same selection.

6. Relevance and Appropriateness--Only those knowledge or performance aspects which are appropriate for the course objectives should be included. Too often evaluation studies focus on the measurement of reader reactions and not on learning or behavior on the job.

Examples of Dependent Variables

In evaluative research, the reader's response is always considered the dependent variable; i.e., that which is observed and measured. The independent variable is the variable that is manipulated to see what effect changes in that variable will have on the dependent variable(s). In a readability study, the dependent variable could be the trainee's responses on an achievement test. An example of an independent variable in a readability study is the variation of the textual materials under test.

In order to obtain a dependent variable measure which reflects student learning on the basis of reading textual material, tests will have to be constructed. To accomplish this, an outline of the written material should be prepared first. The outline need only be in terms of major headings or categories. The test(s) should measure a representative sample of these categories. Decide, in advance of constructing the test(s), on the number of items that should be written in each area. The considerations that go into the selection of the number of items to include in the various areas are: the importance of each area in the total learning picture, the amount of space devoted to each area in the written material, and the objectives that have the greatest value. It is also known that as the number of items in a test increases, the reliability of the test increases.

There are several different types of tests that can be constructed for use in a readability study. Several of the most appropriate are presented below. For a more thorough discussion of these tests, refer to the Handbook for the Evaluation of Technical Training Courses and Students, AFHRL-TR-72-15.

Written Tests

When preparing a written test (also referred to as a paper and pencil test), an item pool of more than the finally required number of items should be developed so that if replacement items are needed for the test, they will be available. There are several types or forms of test items that can be used. The advantages and disadvantages of each item type should be considered as well as the learning objective to which the item is aimed. A test item should measure the learning objective as directly as possible and, depending on purpose, certain item types are better than others.

The multiple-choice item, along with true-false and matching items, are referred to as "selection" items because the student selects his response from alternatives that are offered to him. The selection items have certain advantages over "supply" items (items for which the student has to furnish his own response as in the essay item). The advantages of the selection type item are: (1) test scorer personal biases and subjective idiosyncracies are removed from the scoring, (2) ease of scoring (scoring keys or machines can be used), (3) more aspects of the material can be tested in a given period of time, (4) with increases in comprehensiveness improved validity may be expected, and (5) ease of application of statistical item analytic procedures to resultant data.

Criterion and Norm Referenced Testing

If a criterion referenced test is employed to judge the readability of textual materials, the effects of reading the materials are compared with an absolute standard to determine whether or not acceptable (criterion) performance has been achieved. The absolute standard may be passing a performance item which is at a given level of difficulty or achieving an absolute score. By contrast, the students' results on a norm referenced test are compared with the performance of a peer or reference group; e. g., a given percentile.

The characteristics of criterion referenced tests are that they indicate the degree of competence attained from reading the material without reference to the performance of others, and they measure student performance with regard to specified absolute standards (criteria) of performance.

Generally, criterion referenced tests tell us how effective the written materials are with regard to a specified standard of behavior. Criterion objectives, expressed in terms of specific behaviors, must be identified. Either performance testing or paper and pencil testing can be used with criterion referencing. In criterion referenced measurement, grades are usually S (satisfactory) or U (unsatisfactory).

Performance Tests

In performance tests, the adequacy of the materials being evaluated is tested through the performance of the reader on tasks which are relevant to his present or future job. Performance tests can range from simulated performance through performance of job tasks using actual job equipment. Scoring can be based on measurement of performance in process, adherence of a final product to prescribed standands, care and use of tools during performance, adherence to safety precautions, or some combination of these categories. Completion of a prescribed task within an allotted time is also sometimes scored. One of the most popular methods of scoring is through a sequential checklist. In the sequential checklist, a task is broken down into the sequential elements which must be correctly performed if the task is to be completed. Each of these elements is then sequentially listed for scoring by an examiner while he observes the performance of the person being tested on the task.

The final score is determined by summing the points. A single performance test can be constructed which includes items dealing with safety precautions, quality of the final product, and the manner in which the student performs.

There are variations to performance tests which avoid the need for the student to perform a task. One such variation is the examinee recorded performance test. In this type of performance test, the student is required to write his response to a written question about a picture or diagram of some performance aspect. His response is usually a checkmark or a one-word answer. For example, in one test of this type, the student is asked to identify an error in a picture of someone engaged in an activity similar to that for which the student is training.

The Evaluative Milieu

It is difficult to complete an evaluative study in a training environment which resists change. If the administration perceives revised materials as negatively affecting vested interests, many obstacles can be placed in the evaluator's path. In this case, the evaluator must meet with the persons who are resisting his efforts and attempt to persuade them to the benefits of the evaluation. Often, by making these persons a part of the planning committee for such a study, such negativity can be overcome.

Additional Reading List

Campbell, D., & Stanley, J. Experimental and quasiexperimental designs for research. New York: Rand McNally, 1963.

Cochran, W. G., & Cox, G. R. Experimental designs. New York: Wiley, 1957.

Gronlund, N. Constructing achievement tests. Englewood Cliffs, N.J.: Prentice-Hall, 1968.

Guilford, J. Fundamental statistics in psychology and education. New York: McGraw-Hill, 1965.

Siegel, A. I., Bergman, B. A., Federman, P., & Sellman, W. S. Handbook for the evaluation of technical training courses and students. Lowry Air Force Base, Colorado, AFHRL-TR-72-15, February, 1972. (AD-753 094)

Suchman, E. Evaluative research: Principles and practice in public service and social action programs. New York: Russell Sages Foundation, 1967.

Tyler, R. Education evaluation: New roles, new means. Chicago: University of Chicago Press, 1969.

Wittrock, M., & Wiley, D. (Eds.) The evaluation of instruction. New York: Holt, Rinehart & Winston, 1970.

Wood, D. Test construction. Columbus, Ohio: Merrill, 1960.

REFERENCES

Bormuth, J.R. Comparable cloze and multiple-choice comprehension test scores. Journal of Reading, 1967, 10, 291-299.

Buros, O.K. (Ed.) The Sixth Mental Measurement Yearbook. Highland Park, N.J.: Gryphon Press, 1965.

Caylor, J.S., Sticht, T.G., Fox, L.C., & Ford, J.P. Methodologies for determining reading requirements of military occupational specialties. Presidio of Monterey, Calif.: Human Resources Research Organization, HumRRO Division No. 3, Technical Report 75-5, March 1973.

Dale, E., & Chall, J.S. A formula for predicting readability (and) instructions. Educational Research Bulletin, 1948, 27. 11-20, 28, 37-54.

Dale, E., & Chall, J.S. The concept of readability. Elementary English, 1949, 26, 19-26.

Dale, E., & Tyler, R. W. A study of the factors influencing the difficulty of reading materials for adults of limited reading ability. Library Quarterly, 1934, 4, 384-412.

Federman, P.J., Macpherson, D. H., & Siegel, A. I. Normative development for submarine sonar manuals. Wayne, Pa.: Applied Psychological Services, Inc., 1970.

Flesch, R. Estimating the comprehension difficulty of magazine articles. Journal of General Psychology, 1943, 28, 63-80. (a)

Flesch, R. Marks of readable style. New York: Bureau of Publications. Teachers College, Columbia University, 1943. (b)

Flesch, R. The art of plain talk. New York: Harber & Bros., 1946.

Flesch, R. A new readability yardstick. Journal of Applied Psychology, 1948, 32, 221-233.

Flesch, R. The art of readable writing. New York: Harper, 1949.

Flesch, R. How to test readability. New York: Harper, 1951.

Flesch, R. Reply to "Simplification of Flesch reading ease formula." Journal of Applied Psychology, 1952, 36, 54-55.

Flesch, R. How to make sense. New York: Harper, 1954.

Gray, W.S., & Leary, B.E. What makes a book readable. Chicago: University of Chicago Press, 1935.

Grieme, R., Cleveland, D., Chubb, G., et al. Project PIMO Final Report: Volume IV - PIMO technical data format specifications. TR-69-155, May 1969. Space and Missile Systems Organization, Norton Air Force Base, Calif. (AD-852 104)

Lively, B. A., & Pressey, S. L. A method for measuring the "vocabulary burden" of textbooks. Educational Administration and Supervision, 1923, 9, 389-398.

Lorge, I. Predicting reading difficulty of selections for children. Elementary English Review, 1939, 16, 229-237.

Lorge, I. Predicting readability. Teachers College Record, 1944, 45, 404-419.

Lorge, I. The Lorge and Flesch readability formulae: a correction. School and Society, 1948, 67, 141-142.

Madden, H.L., & Tupes, E.C. Estimating reading ability level from the AQE General Aptitude Index. Lackland Air Force Base, Texas: Personnel Research Laboratory, Aerospace Medical Division, PRL-TR-66-1, AD-632 182, February 1966.

McLaughlin, G. H. SMOG grading: A new readability formula. Journal of Reading, 1969, 12, 639-646.

Ojemann, R. The readability of parents and factors associated with the reading difficulty of parent education materials. In G. D. Stoddard (Ed.), Researches in parent education II. Iowa City, Ia.: University of Iowa, 1934. Cited by J. S. Chall, Readability - an appraisal of research and application. Columbus, Ohio: Ohio State University, Bureau of Educational Research, 1958.

Powers, R. D., Sumner, W. A., & Kearl, B. E. A recalculation of four adult readability formulas. Journal of Educational Psychology, 1958, 49, 99-105.

Smith, E. A., & Sentner, R. J. Automated readability index. Wright-Patterson Air Force Base, Ohio: Aerospace Medical Division, AMRL-TR-66-220, November 1970. (AD-667 273).

Spache, G. A new readability formula for primary grade reading materials. Elementary School Journal, 1953, 53, 410-413.

Taylor, W. Cloze procedure: A new tool for measuring readability. Journalism Quarterly, 1953, 30, 415-433.

Thorndike, E. L. Improving the ability to read. Teachers College Record, 1934, 36, 1-19, 123-144, 229-241.

The integrated maintenance manual. Anaheim, Calif.: Technical Operations, Inc.

Tribe, E. B. A readability formula for the elementary school based upon the Rinsland vocabulary. Unpublished doctoral dissertation, University of Oklahoma, 1956. Cited by G. R. Klare, The measurement of readability. Ames, Ia.: Iowa State University Press, 1963, 69-70.

Vogel, M., & Washburne, C. An objective method of determining grade placement of children's reading material. Elementary School Journal, 1928, 28, 373-381. Cited by C. R. Klare, The measurement of readability, Ames, Ia.: Iowa State University Press, 1963, 38-39.

Washburne, C., & Vogel, M. What books fit what children? School and Society, 1926, 23, 22-24. Cited by G. R. Klare, The measurement of readability, Ames, Ia.: Iowa State University Press, 1963, 38-39.

Wheeler, L. R., & Smith, E. H. A practical readability formula for the classroom teacher in the primary grades. Elementary English, 1954, 31, 397-399.

Wilmot, H. L., Chubb, G. P., Tabachnick, B. J., et al. Project PIMO Final Report: Volume VI - PIMO technical data preparation guidelines. TR-69-155, May 1969. Space and Missile Systems Organization, Norton Air Force Base, Calif. (AD-852-106).

APPENDIX A

How to Perform a Cloze Analysis

The cloze procedure involves: deleting every nth word (usually the fifth word) in passages selected from the text under consideration, administering the mutilated passage to a sample of typical readers and asking them to fill in the blanks, and scoring the answers.

Sampling Reading Passages

Since it will be very difficult and impractical to test all the reading material in a particular manual, book, or course, you will have to resort to a sampling technique. The sampling technique employed should provide samples that are representative of the reading matter and that provide adequate and sufficient data for analysis. To accomplish this, you should:

1. Select samples of approximately 500 words, but not less than 250 words

2. Select samples from each chapter or section (or whatever other divisions are used in the written material). Try to take two samples from each section.

3. Use a random technique for selecting a particular sample. A table of random numbers may be used (these tables are often included in standard statistics text books) such that a digit from the table would identify the paragraph to use in the sample. This would require prenumbering all the paragraphs in the reading matter.

4. Begin each sample at the beginning of the paragraph. Use as much text (250-500 words) as necessary to provide a minimum of 50 blanks per sample.

Preparing the Sample Reading Passages

The words can be deleted manually or by computer program. The computer program would be advantageous if the samples were extensive in length. If alternate forms of the cloze test are needed (e.g., to prevent compromising answers of two groups of subjects who will be tested at different times, if the same reader will be tested at different times but on the same passages, or if retention over successive time periods is of interest) different deletion patterns can be used. This could be accomplished by deleting every fifth word starting with the first, second, third, or fourth word. Other deletion patterns such as deleting every sixth, seventh, or even tenth word could also be used. The five word pattern, starting with the first word in the paragraph is the most popular. The closer the blanks are to each other, the more difficult the task becomes. As the deletion pattern increases from every fifth word, the passage length will necessarily increase in order to have a passage with at least 50 blanks (e.g., if every sixth word is deleted, a 300 word passage will be needed; if every tenth word is deleted, a 500 word passage will be needed). In applying the cloze technique, the following also applies:

1. End each test passage at the end of a sentence. It is not imperative to use the same number of words in each test passage since comparisons will be made on a percentage correct basis.

2. Words are defined as a unit with spaces on either size. Thus, numbers or groups of letters may be considered as words. Examples of single words are: 100, AFB, isn't, on-the-job, and i.e. Punctuation marks such as hyphens, apostrophes, and commas are deleted along with the words.

3. Hyphenated words are not deleted as units if the two parts are not bound to each other, as in on-the-job, but are free forms, as in readability-comprehensibility. In the latter case, the hyphen is not deleted with the word.

4. All blanks should be of the same size. A 15 space blank is most commonly used.

Subjects

In order to be able to generalize the results of the readability study, a sample of subjects must be selected that is representative of the population who normally read the material. Wherever possible, a random sample of the entire population of concern should be selected. This is often impossible to accomplish because the group may be separated geographically, because of time and cost restrictions, or because of other reasons.

Procedures

When administering the reading passages, a full set of directions should be provided to the subjects. A technique often used in this situation is to have the subjects read the directions to themselves while the administrator reads them aloud. The directions should inform the subjects of the purposes of the study, the manner in which the blanks should be filled in, and any restrictions in time limit (although rigid time restrictions should not be used in a cloze analysis). The subjects should be reassured that the study they are participating in involves a test of the material they are reading and not of them. An example passage should be provided. The subjects should be given a few minutes to record their responses in the example and then the correct answers should be read to them. Some other points that may be made in the directions are: (1) enter only one word or number in each blank space, (2) guess if necessary since there is no penalty for guessing, and (3) write down a word even if the spelling is incorrect. After providing the subjects with an opportunity to ask questions, they should be told to begin. A sample set of instructions that was used in one study is presented below.

Instructions

This questionnaire is made up of six reading passages. Some of the passages come from training materials being used now. Others were specially written for comparison. In each passage, about one-fifth of the words have been left out. We want you to write in each blank space the exact word you think is missing.

For example, in the sentence, "George Washington chopped down a cherry tree," the person responding thought that the words "Washington" and "cherry" were the missing words, so he wrote them in the blank spaces.

Some of the missing words will be easy, and others will be quite hard. Sometimes the word will "pop" into your mind; at other times clues in other parts of the passage will help you decide on the right word. In any case, you should try to write in each blank the exact word you think was taken out.

As you finish each section, go right on to the next one. Do not rush through the materials. You will have about an hour and a half to finish, or about 15 minutes for each section.

In scoring responses, accept as correct those responses that are identical to the deleted word. Synonyms are not correct, nor are changes in tense. However, words are accepted as correct even if they are misspelled.

Results

In analyzing the results you should:

1. Tabulate the number of correct responses on each passage

2. Determine each subject's comprehension level by calculating the percentage of correct responses (in the cloze procedure, with every fifth word being deleted, approximately 44 per cent correct is considered acceptable for use

in supervised instruction and 57 per cent correct for use in independent instruction)

3. Assign a reading grade level to each passage (in order to do this, data on many subjects must be obtained and statistical procedures used which would allow for relating the cloze scores of a group of subjects to a reading grade level). The statistical procedures involved are rather sophisticated and therefore reference to a statistical source is recommended. Reading grade levels for the subjects may be obtained by using the Madden and Tupes procedure, as described in Chapter I, Reading Level Determination.

APPENDIX B

Examples of Before and After Material and Automated Readability Index (ARI) and Grade Level (GL) Analyses

The following three examples demonstrated the effect on readability of revised reading material. The excerpts were taken from Air Force manuals, revised in accordance with the principles for making written copy more readable, and then ARI and reading grade levels were calculated.

Before

The receiving inchecker's first duty is to process incoming shipments from the transportation section, commercial carriers, or parcel post. He insures that the total number of boxes, drums, parcels, etc., received are correctly reflected on the receiving documents that accompany the shipments. He also checks for physical damage that may have occurred during transit or as the result of improper packaging or handling by the carrier. He explains noted discrepancies on the reverse side of the receiving document and obtains the signature of the carrier as acknowledgment of the discrepancies. Property is segregated by priority, cost category, and Federal supply class as it is unloaded. If this is not feasible, due to volume or time limitations, segregation is accomplished prior to processing the property through the receiving and inspection lines.

Classified items are not allowed to remain on the receiving line, awaiting unchecking and inspection actions. Upon completion of receiving checkings and inspection action, the receipt is input via the remote keyboard printer.

ARI = 70.50
GL = 15.01

After

The most important job of the receiving inchecker is to process all incoming shipments. Shipments may come from the transportation section, commercial shippers, or parcel post. The inchecker makes sure that the number of boxes, drums, or other containers received is shown on the papers that come with the shipment. He also looks to see if the shipment is damaged. Shipments may be damaged during shipping or because of rough handling or bad packaging. If any of the numbers are wrong or if he finds any damage, he writes it down on the back of the receiving papers. The person who delivered the material must also sign the back of the form. This proves that the delivery man knew that the shipment was damaged, or was too big or too small. While it is being unloaded, the shipment is separated by: (1) priority, (2) cost, and (3) Federal supply class. If the shipment is too big to operate then or if you don't have enough time, it is all right. But, it must be separated before it goes to the receiving and inspection lines. You must not make classified items wait on the receiving line for inchecking and inspection. When the checking and inspecting is done, the receipt of the shipment is input via the remote keyboard printer.

ARI = 58.79
GL = 8.95

Before

Radio interference suppression is the elimination or minimizing of the electrical disturbances which interfere with radio reception, or disclose the location of the vehicle to sensitive electrical detectors. It is important that vehicles with or without radios be properly suppressed to prevent interference with radio reception of neighboring vehicles.

The ignition and generating systems have been designed to suppress radio interference. Ignition system interference suppression is effected by a primary capacitor in the distributor wiring harness receptacle, a resistor suppression in the spark plugs, and by shielded spark plug cables. Radio interference suppression in the generating system is effected by a choke, capacitor, filter, and rectifier in the generator regulator assembly. The primary capacitator is housed in the distributor wiring harness receptacle and is an integral part of the primary circuit. Spark plugs are shielded individually by metallic braid beneath the rubber insulation. The generator regulator assembly houses the capacitor, choke, filter, and rectifier.

ARI = 72.77
GL = 16.73

After

Radio interference is caused by electrical impulses. The interference can be stopped or cut down. Detectors can locate vehicles. All vehicles should be checked so they will not interfere with the radio reception of other vehicles.

Ignition and generating systems were designed so they won't interfere with radios. This is done through 3 parts of the ignition system. One is a capacitor in the distributor wiring harness. The capacitor is part of the primary circuit. Another is a resistor in the spark plugs. The third is the shelded spark plug cables. Each spark plug is shielded by metal braids under rubber insulation. There are 4 items in the generator regulator assembly that cut radio interference. These are the choke, capacitor, filter, and rectifier.

ARI = 56.145
GL = 7.73

Before

The process described for developing the human component of a weapon system has been combined with evolving educational and training methodology to form an effective process for designing and operating training programs--the system process. An instructional system is an integrated combination of resources (students, instructors, materials, equipment and facilities), techniques, and procedures performing efficiently the functions required to achieve specified learning objectives. Under this concept, we must ensure that training programs are not bound by traditional procedures, techniques, or routines of the day. Therefore, instructional system designers must carefully consider both current and future needs when identifying and designing facilities, equipment, and devices. This same total systems concept must be applied when identifying and planning instructional activities. When working in this area, you must develop and organize learning experiences that will enable students to apply what they have learned to new or changing situations. The instructional system must employ teaching techniques that will produce graduates who will not become the victims of fixed habits, procedures, or attitudes, nor be bound by extreme specialization.

ARI = 74.783
GL = 16.79

After

A plan to design and operate training courses was developed. It is known as the system process. It joins the methods used to develop a man in a weapon system with those of education and training. A training plan achieves its learning goals by combining a few things. They are resources, techniques, and procedures. By resources, we mean students, teachers, materials, tools, and facilities. Training courses should not be bound by old fashioned procedures or techniques. Facilities, equipment, and techniques should be designed for current and future needs. Develop and plan the system so that students can apply what they learn to new situations. The techniques used should not make students creatures of habits. Also, students should not be bound by specialization.

ARI = 58.457
GL' = 8.91

APPENDIX C

How to Perform a Flesch Analysis

According to Flesch, readability is composed of two component parts, reading ease (an estimate of the case a reader will have in reading and understanding what was written) and human interest (an estimate of how interesting the reader will find the written material).

Sampling Reading Passages

Since it will be very difficult and impractical to test all the reading material in a particular manual, book, or course, you will have to resort to a sampling technique. The sampling technique employed should provide samples that are representative of the reading matter and that provide adequate and sufficient data for analysis. To accomplish this, the following should be considered:

1. Select three to five samples from an article and 25 to 30 samples from a book.

2. Samples should be taken at random. Tables of random numbers may be used or a scheme such that every fifth paragraph, or every other page, or every tenth page is used. Do not select introductory paragraphs as samples (they are usually not typical of the style of the rest of the section).

3. Each sample should have approximately 100 words.

4. Start each sample at the beginning of the paragraph.

Procedures

The following steps should be performed to determine reading ease:

1. Count each word in the sample, marking off the 100th word.

2. Words are defined as a unit with spaces on either side. Thus, numbers or groups of letters may be considered as words. Examples of single words are: 100, AFB, isn't, on-the-job, and i.e.

3. Calculate the average number of words in each sentence. That is, count all the sentences in the sample. Then, divide the number of words by sentences. For instance, if you have a 110 word passage and 12 sentences, the average sentence length is 9 (rounding off to the nearest whole number). If more than one sample is selected from a piece of writing, add the total number of sentences in all the samples and divide the total number of words by the total number of sentences.

4. If you restrict the sample to exactly 100 words, then the sentence in which the 100th word falls is counted in the sentence count if more than half the words in the sentence were included in the 100 word count. For example, if the last sentence was "I sang for my supper," and the 100th word in the sample was sang, this sentence would not be counted.

5. The following punctuation marks define a sentence: period, question mark, exclamation point, colon, and semicolon. However, sentence counting becomes complicated with colons and semicolons in certain situations. In order for the clauses to be counted as separate sentences, they must be independent and a separate unit of thought. For

example, the following is counted as one sentence: The advantages of this technique are: (1) ease of use; (2) validity. But, the next example is counted as two sentences: He is tall and dark; she is short and blond.

6. Count the average word length in syllables. This is done by counting all syllables and dividing the total number of syllables by the number of words. In the final formula, the average word length in syllables is expressed as the number of syllables per one hundred words. Therefore, multiply the result of the total sample by 100. When using 100 word syllables, count the total number of syllables in all the samples and divide by the number of samples.

7. Syllables are counted the way the word is pronounced. For example, average has two syllables, and readability has five syllables. Syllables, for numbers and symbols, are counted the way they are pronounced; e.g., AFB has three syllables, # (number) has two syllables, and 34 (thirty-four) has three syllables.

8. The reading ease score is determined by the following formula:

   Reading Ease Score = 206.835 - [1.015 (average sentence length) + .846 (number of syllables per one hundred words)]

This result will vary between 0 (unreadable) and 100 (easy to read). The table below may be used to help interpret the reading ease scores.

Reading Ease Score	Interpretation	Grade Level
90-100	Very Easy	5
80- 89	Easy	6
70- 79	Fairly Easy	7
60- 69	Standard	8 and 9
50- 59	Fairly Difficult	10 to 12
30- 49	Difficult	13 to 16
0- 29	Very Difficult	College graduate

To find the human interest score:

1. Count the number of personal words per 100 words. In testing a whole piece of writing, divide the total number of personal words by the total number of words and multiply by 100. If 100 word samples are used, count the personal words in each sample and divide the total number of personal words in all samples by the number of samples.

2. Personal words are the first, second, and third person pronouns. The neuter pronouns it, its, itself are not counted. The pronouns they, them, their, theirs, themselves are counted if they refer to people rather than things. He, him, his, she, her, and hers are always counted (even if they refer to animals or things).

3. Count all words that are of the masculine or feminine form, such as Betty, mother, brother, John, ballerina. However, words that are common to both sexes are not counted; e.g., lawyer, parent, singer, child. Doctor John Smith has one personal word (John); Mr. John Smith has two personal words (Mr. and John).

4. Count group words; e.g., folks and people as personal words.

5. Count the number of personal sentences per 100 sentences. In testing a whole piece of writing, divide the total number of personal sentences by the total number of sentences and multiply by 100. In testing samples, divide the number of personal sentences in all the samples by the number of sentences in all the samples and multiply by 100.

6. Personal sentences are spoken sentences marked by quotation marks or other indications that the sentence was spoken (spoken sentences are sometimes set off by commas and colons). Do not count quoted phrases when the quotes do not indicate spoken phrases; e. g., The Flesch readability technique measures "reading ease" and "human interest."

7. Count as personal sentences all sentences that are addressed to the reader; e. g., Add 3 + 2, Count all personal sentences, Do not count words that are common to both sexes. Do not count sentences that are vaguely addressed to the reader; e. g., It is a good idea to arrange the data neatly.

8. Exclamations are counted as personal sentences; e. g., Holy cow!

9. Incomplete sentences whose meaning has to be inferred are counted as personal sentences; e. g., No, So, So.

10. The human interest score is determined by the following formula:

    3. 635 (number of personal words per 100 words) + . 314 (number of personal sentences per 100 sentences)

    This result will vary between 0 (no human interest) and 100 (much human interest). Using the following table to interpret the human interest score:

Human Interest Score	Interpretation
60 to 100	Dramatic
40 to 59	Highly Interesting
20 to 39	Interesting
10 to 19	Mildly Interesting
0 to 9	Dull

Example

When thinking of "What will the student be able to do?" we must first answer the question "What can the student already do?" This has a direct impact on the analysis of each training standard task/subject knowledge statement and eventually affects the depth and scope of the learning objectives. For example, if the typical student's previous work has involved related tasks, many elementary details may be omitted. This applies to the analysis of the training standard and the development of objectives. AFM 50.5, USAF Formal Schools Catalog and paragraph 3 of the specialty description in AFM 36-1 or AFM.

Reading Ease Score

$$\text{Average sentence length} = \frac{\text{no. of words}}{\text{no. of sentences}} = \frac{100}{5} = 20$$

Number of syllables = 179

$$\text{Reading Ease Score} = 206.835 - [1.015(20) + .846(179)] =$$
$$206.835 - [20.3 + 151.434] =$$
$$206.835 - 171.734 = 34.101$$

Human Interest Score

Personal words = 1
Personal sentences = 1

$$\text{Human Interest Score} = 3.635(1) + .314(1) = 3.949$$

APPENDIX D

How to Perform the FORCAST Analysis

The FORCAST technique provides a reading grade level based on a monosyllabic word count. An example of the technique is demonstrated below. The sample reading passage was excerpted from a military guide for instructors.

> An instructor cannot read the mind of a student to verify the extent of his understanding. It is only through some overt activity on the part of the student that the extent of his knowledge or skill can be determined. To say that a student will "develop an understanding of Ohm's Law" is often used as an example of an obscure objective. One person may feel that if the student writes the law it indicates understanding. Another may say that the student should explain the law. A third may contend that the only way the student can satisfy this objective is for him to use the formula to solve problems in electrical circuitry.

Number of one syllable words in a 150 word passage = 76

$$\text{Reading Grade Level} = 20 - \left(\frac{\text{no. of one syllable words}}{10}\right) = 20 - 7.6$$

$$= 12$$

APPENDIX E

How to Perform the SMOG Analysis

The following example demonstrates the procedure for calculating a SMOG grade on a passage of 30 sentences.

Sample Test

1. GENERAL INFORMATION

The need for human engineering in the national defense is based on the fact that the science of man and his capacities must keep pace with technology if military effectiveness is to be achieved. Machines cannot fight alone; they require men to operate and maintain them. The task of human engineering, both generally and as an element of the Personnel Subsystem (PS), is to get the best from man in the operation of the highly complex equipment produced by modern technology. (Some of the concepts expressed in this and the following paragraphs are based on Ref 193.) Human engineering proceeds on the assumption that the capacities of man develop within certain limits and that by adapting the design of equipment to these capabilities, a more effective system will result. A system that requires extreme, uninterrupted concentration over long periods of time would not be "human engineered" nor could a system be relied upon in combat if it were so complex that it could be maintained only by the engineers and scientists who designed it. As used in this handbook, the term "human engineering" is not synonymous with human factors. The term "human factors" is more comprehensive, covering all biomedical and psychosocial considerations applying to man in the system. It includes not only human engineering, but also life support, personnel selection and training, training equipment, job performance aids, and performance measurement and evaluation. Human factors is closely related to and is often used as a synonum for PS. Sometimes the total human factors effort (including human engineering and other PS elements) is called "bioastronautics" or "life sciences."

2. THE PRACTICE OF HUMAN ENGINEERING

System performance must not be degraded by improper design which imposes limitations on the human operator or maintenance technican. Consequently, the Air Force Systems Command (AFSC) has been assigned responsibility for ensuring that new systems are planned and designed so as to exploit known human capabilities, but, at the same time, not require any performance which exceeds those capabilities. Carrying out this responsibility constitutes the "practice" of human engineering in the Air Force context.

2.1 Who Does Human Engineering?

No single professional group is responsible for research in human engineering or the application of human engineering principles to the design of systems. Engineers, industrial engineers, anthropologists, physicians, physiologists, and psychologists all participate. In a broad sense, everyone concerned with the design or improvement of devices which people use or the ways in which they use them is practicing human engineering. However, as so often happens in a technical field, a point is reached where every man can no longer be his own specialist. The profession of human engineering has emerged to meet this demand. Human engineering draws its basic data from several fields of knowledge, among them (1) the engineering sciences, (2) experimental psychology, (3) physical anthropology, and (4) psychophysiology. The profession calls for varied talents and broad training in both the physical sciences and the sciences that deal with man. It is understandable, therefore, that many human engineering specialists are experimental psychologists conversant with engineering or engineers with supplementary training in the behavioral sciences.

2.2 What Do Human Engineers Do?

Human engineers conduct research, establish principles, and assist in designing and developing system hardware that is properly engineered for effective human use. They study the human being as a user of equipment and equipment as a tool for a task. In collaboration with hardware engineers and designers, human engineers seek to develop new and improved man-equipment interfaces that will simplify the operator's task and increase the probability of mission accomplishment. They seek to achieve displays that will most effectively present information to the human senses, to obtain the most efficient controls for human use, and to provide an optimum work environment. They contribute to the development of such complex man-equipment systems as combat-information centers, air defense systems, space vehicles, and offensive weapon systems. In the design of military equipment, human engineers place emphasis on efficiency, as measured by the speed and accuracy of performance, and on the safety and comfort of the human operator. Because the successful design of equipment for human use requires consideration of man's basic characteristics, human engineers study man's sensory capacities, his muscular strength and coordination, his body dimensions, his perception and judgment, his native skills, his optimum workload, and his requirements for comfort, safety, and freedom from environmental stress.

3. HUMAN ENGINEERING IN THE AIR FORCE

Within the Air Force, AFSC is responsible for both the development and the application of human engineering knowledge and techniques.

Sample Calculation

$$\text{SMOG Grade} = 3 + \sqrt{\text{words of 3 syllables or more in 30 sentences}}$$

$$= 3 + \sqrt{179} = 3 + 13.4$$

$$= 16$$

APPENDIX F

Glossary of Readability Measures

Author(s)	Date	Criterion	Variables	Predicted Value	Formula	Type of Material Studied	Range of Material Studied	Method of comparison with criterion
Lively & Pressey	1923	Judged difficulty	Weighted median index no., based on Thorndike word list	Relative difficulty	Weighted median index number	School, general, scientific	Grade 2-college	Inspection
Dolch	1928	Grade level assigned by publishers	Five indices of vocabulary "load	———	Base judgment of difficulty on "several of vocabulary factors	Reading texts	Primer-grade 4	Inspection
Washburne & Vogel (Winnetka)	1928	Mean tested reading level of those liking book	X_2: Number of different words in 1000 words X_3: Number of prepositions in 1000 words X_4: Number of words not on Thorndike list of 10,000 words X_5: Number of simple sentences in 75 sentences	Reading score on Stanford Achievement Test (X_1)	$X_1 = .085X_2 + .101X_3 + .604X_4 - .411X_5 + 17.43$	Children's library books	Grades 3-9	Correlation .845 (This value is confounded by effects of popularity of books)
Lewerenz	1929	Order of presentation on Stanford Achievement Test	Percentage of words beginning with W, H, B (easy) I, and E (hard)	Grade level	Mean of tabled values for each of variables	Stanford Achievement Test paragraphs	Grade 2-college	Graphic
Lewerenz	1930 1935 1939	———	1) Ratio of Anglo-Saxon to Greek and Roman words (difficulty) 2) Ratio of words in "Clark's first 500" to total different words (diversity) 3) Estimate of image bearing or sensory words (interest) 4) Polysyllabic word count 5) Vocabulary mass (number of different words in sample that are not on Clark's list of 500 common words	Grade level	Mean of tabled values for each of variables	———	———	———
Johnson	1930	Grade level assigned by publishers	Percentage of polysyllable words	Grade level	Tabled value of percentage of polysyllables	Wide variety	Primer to grade 8	Inspection
Patty & Painter	1931	None (relied on validity of Thorndike word list)	1) Thorndike word list index numbers 2) Number of different words in sample	Relative difficulty	Index Number = $\frac{A.W.W.V.}{12}$ A.W.W.V. = mean Thorndike Index Number R = number of different words in sample	School texts	Grades 9-12	———

BEST COPY AVAILABLE

Author(s)	Date	Criterion	Variables	Predicted Value	Formula	Type of Material Studied	Range of Material Studied	Method of comparison with criterion
Olemann	1934	X_{C50}	6 measures of vocabulary difficulty, 8 measures of structure, 4 qualitative factors	———	16 passages are presented in order of difficulty, to be compared with material to be tested	Adult education	Grade 6-college	Quantitative factors: correlation. Individual factors up to .60 Qualitative factors: inspection.
Dale & Tyler	1934	Comprehension score of adults of low reading level	X_2: Number of different technical terms X_3: Number of different hard non-technical words X_4: Number of indeterminate clauses	Percent of adults of low RGL (3-5) who will understand passage (X_1)	$X_1 = -9.4X_2 - .4X_3 + 2.2X_4 + 114.4 \pm 9.0$	Adult health education material	Below grade 8	Correlation .511
McClusky	1934	Reading speed, comprehension score	Vocabulary, sentence structure	———	No formula. Description of vocabulary and sentence structure characteristics in easy and hard materials.	Wide variety	Above grade 8	Inspection
Thorndike	1934	Vocabulary difficulty	Number of words not on Thorndike list in 10,000 word sample	Grade level	Tabled value of number of words not on Thorndike list.	General fiction and non-fiction	Grade 4-9	———
Gray and Leary	1935	Mean comprehension test score of adults	X_2: Number of different words not on Dale List of 769 X_5: Number of personal pronouns X_6: Mean sentence length in words X_7: Percent different words X_8: Number of prepositional phrases	Mean comprehension test score of adults (X_1)	$X_1 = .01029X_2 + .009012X_5 - .02094X_6 - .03313X_7 - .01485X_8 + 3.774$	General adult	Grade 2-college	Correlation .6435
Washburne & Morphett (revision of Vogel & Washburne, 1928)	1938	Mean tested reading level of those liking book. Also teacher judgment at grades 1-2.	X_2: Number of different words per 1000 words X_3: Number of different words per 1000 not in Thorndike's most common 1500 words X_4: Number of simple sentences in 75 sentences	Grade level (X_1)	$X_1 = .00255X_2 + .0458X_3 - .0307X_4 + 1.294$	Children's library books	Grades 1-9	Correlation .86 (This value is confounded by effects of popularity of books)

BEST COPY AVAILABLE

Author(s)	Date	Criterion	Variables	Predicted Value	Formula	Type of Material Studied	Range of Material Studied	Method of comparison with criterion
Stone	1938	———	Ratio of new words to total words, percent sentences complete on one line	Relative difficulty	Relative difficulty based on variables stated.	Reading texts	Grade 1	———
DeLong	1938	———	Percentage of words in various levels of Stone's word list.	Six levels of relative difficulty	Tabled value of percentages of words at various levels of Stone's word list	Reading texts	Pre-primer to grade 2	———
Lorge	1939 1944 1948	McCall-Crabbs test norms	X_2: Percent of words outside Dale list of 769 words X_6: Mean sentence length X_8: Percent prepositional phrases	Reading grade level needed to answer 50% of questions correctly (C_{50})	$C_{50} = .10X_2 + .06X_6 + .10X_8 + 1.99$	McCall-Crabbs test passages	Grades 3-12	Correlation .67
Morris & Halverson	1938	———	4 Qualitative vocabulary classes: I) words learned early in life II) localisms III) concrete words IV) abstract words	Relative difficulty	Tabled norms relate count of words in 4 classes to difficulty	———	———	Multiple correlation of .74 was found between classes I, III, IV, and McCall-Crabbs test norms (Lorge, 1939)
Yoakam	1939	———	Vocabulary index: based on Thorndike word list. Words between 4th and 20th thousand, index is the number of thousand in which word appears. Above 20th thousand index = 20.	Grade level	Mean index number comparison with tabulated values	School texts	2nd -14th grade	Inspection
Kessler	1941	Judged difficulty	Mean sentence length in words Mean number of different hard words per 100 words	Grade level	Factors compared to Gray and Leary standards independently	35 textbooks	Grade 2 - college	Inspection
Flesch	1943	Judgment and McCall-Crabbs test norms	X_s: Mean sentence length X_m: Number of affixes per 100 words X_h: Number of personal references per 100 words	Reading grade level needed to answer 75% of questions correctly (C_{75}) (correction for high grade levels)	$C_{75} = .1338X_s + .0645X_m - .0659X_h + 4.2498$ Later corrected to; $C_{75} = .07X_m + .07X_s - .05X_h + 3.27$	Adult magazines and McCall-Crabbs passages	Grades 3-12	Correlation .74

BEST COPY AVAILABLE

Author(s)	Date	Criterion	Variables	Predicted Value	Formula	Type of Material Studied	Range of Material Studied	Method of comparison with criterion
Flesch (revision)	1948	Judgment and McCall-Crabbs test norms	wl = Mean word length (syllables per 100 words) sl = Mean sentence length pw = Number of personal words per 100 words ps = Number of personal sentences per 100 sentences	Relative difficulty, relative interest	Reading Ease = $206.835 - .846_{wl} - 1.015_{sl}$ Human Interest = $3.635_{pw} + .314_{ps}$	McCall-Crabbs passages	Grades 3-12	Correlation R.E.: .70 H.I.: .43
Dale & Chall	1948	McCall-Crabbs test norms, comprehension test scores	X_1: Percent of words not on Dale list of 3000 common words X_2: Mean sentence length in words	Reading grade level needed to answer 50% of questions correctly (C_{50})	$C_{50} = .1579X_1 + .0496X_2 + 3.6365$	McCall-Crabbs passages, school material, health articles	Grades 3-12, 3-16 for health passages	Correlation .70
Dolch	1948	Grade level assigned by publishers	X_1: Median sentence length X_2: 90th percentile sentence length X_3: Percent hard words (not on Dolch list of 1000 words)	Grade level	Look up three variables in tables, average the tabled values.	Reading texts	Grades 1-6	Inspection
Wheeler & Wheeler	1948	None (relied on validity of Thorndike word list)	Percent of words in each thousand of Thorndike 20,000 word list	Instructional, Independent, and average grade level	Tabulate words in passages by Thorndike thousands. For instructional level, 90% of words must be below student's RGL. For independent level 95%.	———	———	———
Flesch	1950	Comprehension test scores	wl: Word length (syllables per 100 words) dw: percentage of "definite words."	R: Level of abstraction" on arbitrary scale	$R = 168.095 + .532_{dw} - .811_{wl}$	School texts	Grades 3-12	Correlation .72
Farr, Jenkins, & Patterson	1951	Flesch reading ease score	nosw: Number of one-syllable words sl: Mean sentence length	Relative difficulty	New Reading Ease Index = $1.599_{nosw} - 1.015_{sl} - 31.517$	Adult, public relations	Full range of Flesch "reading ease" values	Correlation .95
Gunning	1952	Flesch scores, McCall-Crabbs norms, and judgment	Mean sentence length, percent of words of 3 or more syllables	Fog Index (the reading grade level needed to read and understand material)	Fog Index = .4 x (mean sentence length + percent of words of 3 or more syllables)	General adult and school	Grades 6-12	Inspection

BEST COPY AVAILABLE

Author(s)	Date	Criterion	Variables	Predicted Value	Formula	Type of Material Studied	Range of Material Studied	Method of comparison with criterion
McElroy (AF Manual 11-3)	1953	_____	Easy words (pronounced in 1-2 words), assigned value 1 hard words (pronounced in 3 or more syllables), assigned value of 3	Fog Count (arbitrary value) and reading grade level	Fog Count per sentence = sum of 1's and 3's in sentence RGL = sum of 1's and 3's divided by number of sentences. If value is over 20, divide by 2. If below 20, subtract 2, then divide by 2.	_____	_____	_____
Forbes & Cottle	1953	Mean value of five existing readability formulas	Mean Thorndike vocabulary difficulty index, words in 4th thousand and above.	Reading grade level	Look up mean vocabulary index in tables to obtain RGL	Standardized tests	Grade 5-college	Correlation .96
Spache	1953	Grade level assigned by publishers	X_1: Mean sentence length X_2: Percentage of words not on Dale list of 769 easy words	Grade level	Grade level = $.141X_1 + .086X_2 + .839$	Grade school tests	Grades 1-3	Correlation .82
Flesch	1954	Observed characteristics of easy and hard material	- Number of references to human beings, etc., per 100 words - Number of indications of voice communication per 100 words.	Arbitrary scales of realism ("r") and energy ("e")	"r" score: tabled value corresponding to observed number of references to human beings, etc. "e" score: tabled value corresponding to observed number of indications of voice communication	General adult material	Adult	Inspection
Wheeler & Smith	1954	Grade level assigned by publishers	Mean length of "units" (similar to sentences), percentage of multi-syllable words.	Reading grade level	Index = 10 (mean length of units x percent multi-syllable words). Look up index in table to obtain RGL	School texts	Grades 1-4	Inspection
Tribe	1956	McCall-Crabbs test scores (revised form)	X_1: Mean sentence length X_5: Percentage of words not on Rinsland word list, multiplied by 100.	Reading grade level needed to answer 50% of questions correctly (C_{50})	$C_{50} = .0719X_1 + .1043X_5 + 2.9347$ -Correction is then applied through reference to table of Dale and Chall (1948)	McCall-Crabbs passages, 1950 revision	Grades 3-12	_____
Gillie	1957	Flesch "level of abstraction" score.	X_1: Number of finite verbs per 200 words X_2: Number of definite articles with nouns per 200 words X_3: Number of nouns of abstraction per 200 words.	Abstraction level (arbitrary scale)	Abstraction level = $36 + X_2 + X_1 - (2X_3)$	School texts, general adult material	Grade 4-college	Correlation .83

BEST COPY AVAILABLE

Author(s)	Date	Criterion	Variables	Predicted Value	Formula	Type of Material Studied	Range of Material Studied	Method of comparison with criterion
Coleman (described by Szalay 1965)	1965	Cloze score	X_1: percentage of one-syllable words X_2: Number of sentences per 100 words X_3: Number of pronouns per 100 words X_4: Number of prepositions (not including "to") per 100 words.	X^1: Cloze score	a) $X^1 = -38.45 + 1.29X_1$ b) $X^1 = -37.95 + 1.16X_1 + 1.48X_2$ c) $X^1 = -34.02 + 1.07X_1 + 1.18X_2 + .76X_3$ d) $X^1 = -26.01 + 1.04X_1 + 1.06X_2 + .56X_3 - .36X_4$	Not stated	Grade 2-professional	Correlation a) .86 b) .88 c) .88 d) .90
Smith & Senter (ARI)	1966	Grade level assigned by publishers	w/s: Mean sentence length in words s/w: Mean word length in strokes (letters)	Arbitrary value and grade level	Automated Readability Index = (w/s) + 9(s/w) Grade Level = 0.50(w/s) + 4.71(s/w) -21.43	School texts	Grades 1-7	Correlation (For grade level only) .98
Fry	1968	Grade level assigned by publishers	Mean number of sentences per 100 words Mean number of syllables per 100 words	Grade level	Plot observed values of the stated variables on the graph supplied. Then read off grade level.	School texts	Grades 1-college	Inspection. Correlations with other formulas and student comprehension test score ranged from .78 - .98
McLaughlin	1969	McCall-Crabbs test. (100% score required) (1961 revision)	Number of words of three or more words in 30 sentences	Grade level	SMOG Grade = 3 + square root of polysyllable count	McCall-Crabbs passages	Grades 1-professional	Correlation approximately .7
Coke & Rothkopf	1970	Flesch "Reading Ease" scores	- Words per sentence - Vowels per word (an index of syllables per word)	Flesch "reading ease" scores	Not stated. Contained in computer program.	General adult and McCall-Crabbs passages	Not stated. Probably low grades through adult.	Correlation .92
Caylor, Sticht, Fox, & Ford	1972	Cloze score	Number of one-syllable words per 150 word sample	Reading grade level required for 35% cloze score	$RGL = 20 - \left(\frac{\text{number of 1-syllable words}}{10}\right)$	Army technical literature	Grade levels 6-13	Correlation .77

BEST COPY AVAILABLE